

# Making British justice fair and sure

**Britain's first Royal Commission in a decade is not destined to blunt criticism of British justice, but the importance it attaches to forensic evidence may point to a more effective remedy.**

Two weeks ago in Britain, a male student from Northern Ireland was brought before a British court on terrorism charges, but promptly discharged because the prosecution had no evidence to offer. The student had been arrested three months earlier, after an explosion in the City of London caused by the Irish Republican Army, and had spent the interval in custody. After his discharge by the court, he was promptly rearrested and deported to his home city of Londonderry, which is in the United Kingdom, under the provisions of the Prevention of Terrorism Act. The act does not require that exclusion from the rest of Britain should be justified; indeed, the Home Secretary has said that he has evidence that the student has links with terrorism, but that he does not intend to make it public.

That sounds like injustice. So it is, especially when some thought is given to the possibility that a person identified and blamed from on high will quickly become a marked man to those in Northern Ireland who think it proper to substitute guns for the paraphernalia of the law. It is therefore all the more ironical that this incident should have preceded by only a few days the report of the Royal Commission on Criminal Justice under Lord Runciman (a judge), which was itself created as a response to a string of acknowledged miscarriages of justice linked with terrorism in the United Kingdom (see pp 178 & 179) in the 1970s. The common thread in these affairs was the eagerness of the prosecution to convict. On at least one occasion, unsafe forensic evidence was offered by the prosecution to the court, and not withdrawn when, by the time of an appeal against the conviction, the prosecution must have known it was unreliable.

## Chastened

Inevitably, what the Runciman commission has to say about the future of Britain's forensic service, and on the admissibility of forensic evidence in the courts, cannot be taken in isolation, but will be part of a wider debate. Mercifully, that seems already to have begun. The British, only recently boastful of the quality of their justice, seem properly chastened by its recent failures. That, no doubt, is why most of those who have spoken about Runciman have protested at the recommendation that a defendant's right to trial by jury should be curtailed, and that the defence should furnish the prosecution in a trial with advance warning of its case. Perhaps infected

by the public penny-pinching now again under way in Britain, the commission seems to have been more concerned with the efficiency than the quality of justice. That, it may be thought, defeats the purpose for which it was established.

The forensic issue nevertheless deserves close attention. It is not an accident that similar questions have recently been raised before the Supreme Court in the United States (see *Nature* 364, 94; 1993). Everywhere, judicial authorities are overwhelmed by questions that cannot be answered except by scientific arguments. And they are bamboozled because, on many occasions, the answers to technical questions are not unambiguous, but often are assessments of the relative probability of black on the one hand and white on the other. The US Supreme Court has decided wisely in moving away from the properly maligned Frye rule, whose effect is to endow judgements on scientific matters with an often spurious sense of certainty; it remains to be seen whether the lower courts in the United States will be as pleased that they will now have to accommodate legitimate conflicts of opinion.

## Difficulties

In Britain, which also has an adversarial system of justice, the difficulties are more particular and in many ways more immediate. The use of forensic evidence illustrates more general questions. Under present rules, the prosecution in, say, a criminal trial must tell the defence in advance of the evidence on which it plans to make a case against a defendant. If there are fingerprints, it must say so. ("We found his dabs on the kitchen window", for example.) If there are DNA profiles, it must also say so, and describe what inferences it plans to draw from them. But there are obvious problems of authenticity and of access. There may be occasions when the defence has grounds for questioning the integrity of materials collected for analysis by those (the police) with a legitimate vested interest (under the adversarial system) in conviction. And given that most forensic services in Britain are almost exclusively operated by public authorities, how can defence lawyers easily arrange for independent evaluation of a prosecution's evidence?

Runciman's solution is to resist the government's plan to make the Forensic Science Service an independent entity (to "privatize" it), but to allow that it should be prepared to advise defendants and their lawyers. Specifically, the



commission foresees a network of equipotent laboratories, one working for the prosecution and one for the defence. That could work, but probably not well. Especially in technically novel fields, some laboratories will be more equal than others. And it is unlikely that even the best-provided forensic service would be so well-supplied with expertise that it could advise both sides in a case impartially, and without conflict of interest. And while the reasoning is clear behind the proposal that prosecution and defence should agree before a trial what the forensic evidence is and what interpretations they will each put upon it, impartiality is not self-evident.

That is why it would be far preferable that the forensic services should be neither private nor a creature of the government, but servants of the courts. If it were decreed that, in a criminal trial, forensic evidence would ordinarily be provided by a laboratory functioning under the direct supervision of the judicial system, doubts about the authenticity of evidence could hardly arise. The obvious difficulty is that opportunities would persist for analysis and interpretation of data to rest with individuals (one of the origins of many recent miscarriages of justice), but, as the newest graduate student knows, the best way to make sense of data is to discuss them in an open forum. That argues for a forensic service run on collegiate, not hierarchical, lines. If that should be judged unworkable, the only feasible alternative is to abandon the adversarial system, to precede each accusation by a judicial assessment of the evidence (as by an investigatory magistrate) and to allow judges to employ experts to advise them on the details of technical cases (which already happens in British commercial courts).

Although lawyers may dismiss the notion that the need to deal with forensic and other technical evidence may in itself force a reconstruction of the basis of the British judicial system, that would be mistaken and complacent. Faulty forensic evidence has already been responsible for more than its fair share of faulty convictions, and under the present British system can only become a greater nuisance. So too will multiply the occasions when disputes that can be settled only by an assessment of the probabilities of contradictory events are, instead, decided by the intellectual equivalent of the tossing of a coin, and by people who are not educated to understand the processes involved.

Remedies of that kind will not avoid abuses of administrative law, such as the banishment of students to Northern Ireland for no better reason than that a government minister has been persuaded by untested and unpublished evidence — and that they have Irish accents. And the Prevention of Terrorism Act is justifiable by the continuing (even worsening) violence in Northern Ireland. But those now faced with the need for the dramatic improvement of British justice may reasonably calculate that the present violence owes at least some of its bitterness to the way in which wrongly chosen Irish men and women have been the most prominent victims of injustice in the past few years. □

## Social science's summit

**The Tokyo summit may have put social science on the map, but its problems and their remedies still lie ahead.**

It must mean something good for social scientists that last week's summit meeting of the heads of government of the seven richest countries, called G7, should have spent most of its time on issues with their roots in these recently neglected fields of study. Certainly it is a long time since such a meeting has drawn attention to the increasing age of its participants' populations, and to have given its opinion that steps would have to be taken to remove part of the cost thereof from public budgets. Demography and welfare studies refer. And, for good reason, world leaders are worried about unemployment. Will they now listen more readily to the solutions canvassed for many years by the softer sciences?

The capacity of those concerned to deal effectively with these issues nevertheless depends on the meeting preceding the summit at which Canada, the European Communities, Japan and the United States hammered out the bare bones of an agreement to reduce the tariffs they now levy on imports. If the agreement sticks (it requires ratification in each case) those concerned can expect to become richer, they can expect that the latest version of the GATT, or the General Agreement on Tariffs and Trade (already two and a half years late), may be signed before the year is up and they can expect that the problems that perplexed them last week will be less intractable.

In rich countries such as those of G7, employment should be no problem. Everybody knows the solution. The first need is, by education, to increase the degree and the amount of skill deployed so that technically advanced occupations substitute for those adding very little to the value of the raw materials they consume; then there is enough cash in circulation to employ more of each population in the service occupations (education included). Among G7 countries, France has most consistently followed this path in the past decade; it is to be hoped that its new government does not lose patience while waiting for the rewards to drop into its lap.

The ageing issue is linked with that of employment. For one thing, why set the elderly to growing roses when, healthier than they used to be, they could continue to earn a living for themselves? The now common riposte — that younger people need the jobs more urgently, and that the elderly stultify the organizations to which they belong — has some force, but the less in prosperous societies. But what should be clear everywhere is that the cost of pensions for the elderly should not be counted as a return of savings made over a life at work, but as the cost of welfare for those in need.

Apparent problems are thus potentially illusions. How to make them vanish? Clinching a new GATT is the simplest way. Then there would be funds with which to embark on the soluble problems. But they will not be spent in that way unless heads of government are more resolute than they seemed last week. In this new competition for prosperity, the winners will be found to have looked ahead. □



# Clinton's technology policy under dual siege in Congress

**Washington.** President Bill Clinton's plan to boost US industrial research and development was caught in congressional crossfire last week, and may have been seriously damaged in the process. The president hopes to quadruple the budget of the National Institute of Standards and Technology (NIST) to \$1.4 billion a year by 1997.

Last week, \$220 million of 1994 funding for NIST, including all of its planned construction projects as well as the research money it grants to industrial companies under its Advanced Technology Program (ATP), was knocked out of the Commerce, State and Justice appropriations bill on a technicality.

Although supporters of NIST are confident that most of the money will be restored, the setback highlights the difficulties the administration is encountering in attempting to find large amounts of new money to pay for its technology policy.

NIST, which employs 3,200 people on two major sites at Gaithersburg, Maryland, and Boulder, Colorado, has expertise in physics and engineering. The new administration identified it in February as a promising basis for a major and, in the United States, unprecedented government programme to bolster industrial research, especially in small companies.

But the NIST money was a sitting target for its critics last Thursday because the bill to authorize it had not yet been passed. Any congressman could therefore strike it out by raising a point of order.

Robert Walker (Republican, Pennsylvania), the senior Republican on the House of Representatives Science, Space and Technology Committee and an ideological opponent of the use of state funds for schemes such as the ATP, duly stood up to quash the appropriation. The commerce appropriations subcommittee had proposed spending \$129 million on the ATP and \$30 million on the Manufacturing Extension Partnership, another industry support scheme.

Then Lincoln Diaz-Balart (Republican, Florida), whose interest in science policy has hitherto been inconspicuous, knocked out NIST's \$61-million building programme. Diaz-Balart had noticed that the money would be spent renovating NIST's laboratory at Boulder, Colorado, in the district of David Skaggs (Democrat, Colorado). But Skaggs had earlier supported the closure of Television Marti, a station broadcasting to a very small audience in Cuba, and whose broadcasts are always jammed.

Diaz-Balart wanted revenge on Skaggs, and so struck out the NIST money. Procedural skirmishes of this kind, common on school playgrounds, are not unprecedented in the Congress.

Supporters of the administration's technology policy hope to restore NIST's funding when the House and Senate seek to reach a compromise in the autumn. But this may be more difficult than they expect, given



**NIST's director Arati Prabhakar: will she see her budget increase?**

competing demands for available funds. After the NIST money was struck out, the House voted to spend an extra \$60 million from the same pot on 600 extra guards for the Mexican border — a more pressing concern, to many congressmen, than developing long-term mechanisms for technology transfer.

The ATP, which NIST started three years ago, makes grants of the order of \$1 million to help companies or groups of companies on specific research projects. Clinton wanted an extra \$103 million for ATP this year as part of his economic stimulus package, which fell in Congress in the spring. Next year's budget request for \$200 million had been cut back to \$129 million by the appropriations subcommittee even before last Thursday. All that makes the administration's stated objective of spending \$750 million a year on ATP by 1997 look somewhat ambitious.

Arati Prabhakar — appointed NIST director in May — remains confident about

her agency's future role at the heart of the administration's technology policy. "I'm still optimistic we are going to make some progress", she says. "The appropriations subcommittee approved its biggest ever increase for NIST, and the floor action is really a technicality." Prabhakar also contends that resistance to technology programmes has historically come from the administration, not from Congress.

The objective of ATP's expansion is that it should grow so as to make a substantial impact on US competitiveness, but Prabhakar declines to say just how big it would have to be for that to happen. The administration reckons that, if the United States spends \$150 billion a year on research and development, a federal stimulus of half a per cent — \$750 million — is needed, but "there is nothing magic about that number", Prabhakar says.

US industry is pleased to get what it can from the ATP's coffers, but is not yet convinced that the scheme has justified the faith the administration places in it. "We support the ATP as an experiment, and it was a mistake of the Bush administration to starve it of funds", says Bill Morin of the National Association of Manufacturers. But the industry lobby group does not expect ATP or other NIST programmes to grow as spectacularly as Clinton wants. Morin believes that the idea that NIST would grow that rapidly was "probably unrealistic". "Gridlock is alive and well, and living in Congress."

But if Clinton's technology policy plan is stymied by Congress so that its impact on the \$150-billion research effort will be undetectable, what it is the point of doing it at all? That is just the kind of question that earlier this year killed two of Clinton's main policy objectives — the stimulus package and the energy tax. In both cases, ambitious schemes were whittled down to appease opponents, and then defeated because they looked inadequate to meet their respective objectives.

Henry Kelly, a technology adviser at the Office of Science and Technology Policy in the White House, acknowledges that there is an uphill struggle ahead. "One of our problems is getting people to understand the role a technology policy can play in the economy," he says. "We are committed to making NIST a major player. Obviously it would be much easier to achieve what we want to if the budget was expanding. Given the reality of the budget situation, it is going to be difficult."

**Colin Macilwain**



## ESO's 10 per cent solution for Chile's astronomers

**Munich.** Much sooner than expected, Chilean astronomers have won their claims for more observing time on the European Southern Observatory (ESO) telescopes at La Silla in Chile. They are also likely to win better pay and working conditions.

Faced with a serious threat to plans for the DM570 million (US\$350 million) Very Large Telescope (VLT) at a second site in northern Chile, ESO has comprehensively backed down from its intransigent position based on a treaty signed nearly 20 years ago.

For the past few years, astronomers in Chile have been demanding guaranteed observation time in recognition of their role in hosting the telescopes. They also wanted pay parity with European scientists and technicians.

Sensitive to what it considered imperialistic attitudes on the part of the Europeans, the Chilean government backed the demands, and early this year called for ESO to renegotiate its 1964 contract. One counter in the discussion was the threat that Chile would withdraw cooperation over the 700 square km VLT site at Mount Paranal (see *Nature* **363**, 384; 1993).

Although Chile had set a deadline at the beginning of September, four days of negotiations at the end of June have already produced an agreement that satisfies both sides. Chile has in principle won all its points, but ESO is relieved and happy to have, in exchange, a guaranteed future for its VLT project.

Provided that proposals are assessed as scientifically acceptable, scientists associated with Chilean institutes will now be guaranteed 10 per cent of observing time (what they had asked for) at each of the ESO telescopes at La Silla. Observing time is a valuable currency among astronomers, and there is a hefty waiting list; the telescopes are, on average, three times oversubscribed.

One side-effect of the agreement is that it is likely to encourage European astronomers to form scientific collaborations with Chileans to try to jump the queue, which ESO's spokesman Richard West says may be "no bad thing". The agreement also allocates 5 per cent of viewing time of the VLT to Chilean institutes.

The Chilean government says it will also play its part in helping to develop astronomy in Chile by giving more financial support to the training of Chilean astronomers and to astronomy research. ESO's major argument against giving guaranteed observing time was that there were simply not enough highly trained Chileans to make use of it.

The new agreements are expected to be rubber-stamped at ESO's next council meeting in December.

**Alison Abbott**

## Science faces new treatment in British courts

**London.** A radical overhaul of the way that science is used in British judicial proceedings has been proposed by the Royal Commission on Criminal Justice, which last week published the results of a two-year inquiry into the legal system in England and Wales.

The commission was set up by the government after several convictions by British courts in the 1970s were judged by the appeal court to be "unsafe", on some occasions because the convictions were based on forensic evidence later acknowledged to be flawed. That explains the attention paid by the commission to bridging the gulf between scientific and legal approaches to questions about the validity of evidence.

In particular, the commission suggests that where forensic science experts are due to appear for both sides of a disputed case, they should meet before the trial — if necessary in a formal hearing presided over by a judge — to work out points of agreement and disagreement. The goal would be to reduce the complexities of scientific arguments subsequently laid out before a jury.

Experts appearing in the witness box should not, as at present, be confined to answering questions put to them by lawyers for either side, says the commission. Rather, they should be given the opportunity to volunteer additional information they consider appropriate to the case being tried, a response to long-standing complaints by forensic scientists that their skills are often badly used in court proceedings.

A new Forensic Science Advisory Council should also be set up. This would have broad-ranging responsibilities, including the supervision of the quality of forensic investigations by both public and private laboratories and the encouragement of university-based courses and research programmes.

But the commission throws cold water on the idea that the Home Office's Forensic Science Service (FSS), the main source of forensic evidence in court cases, should be hived off as an independent commercial organization — an avenue currently being explored by the government following its recent white paper on science. The commission says there is no evidence that the FSS would operate more effectively if it were cut off entirely from the Home Office, a conclusion welcomed by scientists working for the FSS, who fear that privatization could threaten their reputation for impartiality.

The commission acknowledges that the FSS has now corrected the shortcomings that led to a number of notorious miscarriages of justice in the 1970s. (One of these was the conviction of six Irishmen for allegedly bombing a pub in Birmingham, largely on the basis of scientific evidence, later

discredited, that two of them had been handling nitroglycerine; the chromatographic test developed for the purpose proved less specific than supposed.)

But the 11 members of the commission suggest that further changes are still needed, given the number of cases in which scientific evidence plays a key role, as well as evidence that a substantial proportion of jury members (10 per cent in one survey) find scientific arguments difficult to follow.

This general move to tighten up the quality and use of scientific evidence has been widely welcomed in the legal community (in contrast to other more controversial proposals in the report, such as limiting a defendant's right to trial by jury). Indeed many of its specific proposals — for example that public sector forensic science laboratories should be made more accessible to defence lawyers — echo suggestions made earlier this year by the House of Lords Select Committee on Science and Technology.

But concerns still remain. Some lawyers argue that the proposals do not go far enough, and that the commission could have taken bolder steps to make forensic science expertise readily available to the defence, for example by recommending the creation of independent forensic science centres with this explicit responsibility.

Others warn that pre-trial hearings, by disclosing to the prosecution the detailed arguments likely to be used in challenging scientific evidence, will require defence lawyers to alert the prosecution to their line of attack — and thus reduce the chances of this attack succeeding in front of a jury.

And there still remains the question of how the government will respond to the commission's proposals, in particular those requiring the Treasury to provide extra funding. The royal commission, for example, backs the suggestion of the House of Lords committee that universities should be encouraged to carry out more forensic science research. It refers explicitly to the contribution of DNA profiling, and suggests that "a number of important forensic problems may be amenable to research, which might in consequence narrow the range of possible disputes between prosecution and defence".

In its earlier reply to the House of Lords, however, the government rejected a suggestion that it should provide more money earmarked for this purpose. Arguing that forensic science is not a scientific discipline in itself, but draws on developments in many areas of research, it suggested that applied research was best undertaken in organizations actively engaged in forensic work, and "funded by those who benefit directly".

**David Dickson**

# DNA database proposal gets cautious welcome in UK

**London.** The technique of DNA profiling, invented less than ten years ago by Alec Jeffreys, professor of genetics at the University of Leicester in Britain, passed a new milestone last week when the Royal Commission on Criminal Justice (see opposite) advocated a national database containing the DNA profiles of all individuals convicted of serious criminal offences, from burglary to murder.

The proposal has in principle been welcomed by virtually all participants in the British legal system, implicitly acknowledging the power of genetic information to identify criminal suspects. But several groups have raised concerns about the operation of a database, suggesting that framing legislation to put the commission's proposals into effect may be time-consuming.

There are two separate proposals in the air. One is that the police be given powers to retain information obtained from DNA samples from anyone convicted of a serious criminal offence, regardless of whether the DNA evidence is relevant to that offence. Their powers would be extended to, for example, the right to take saliva samples without consent.

The second proposal is that DNA data on individuals arrested but found innocent should also be retained, to enable the police to build up statistical information about the frequency with which particular characteristics appear in the population. But the commission stresses that responsibility for keeping this information, which would be stored anonymously, should be given to an organization entirely independent of the police.

Police forces in England and Wales, which are already building up their own collections of DNA samples, have welcomed the commission's proposal for detailed legislation to set up both types of database.

Scientists involved in DNA profiling, however, point out that rapid evolution of profiling techniques means that careful consideration needs to be given to the most cost-effective way of analysing and storing data in a form that will remain compatible with data generated by new analytical techniques as these come along (as well as with foreign databases). "The technology has not settled down yet," says Jeffreys. "It is too early to say what is the right set of markers or the right testing system to have a database which can be used indefinitely into the future."

Perhaps the strongest objections have come from civil liberties groups, based not on criticism of the science involved, but on the uncertainties that still surround both the recognition of individual DNA profiles and interpretation of their statistical significance. Given such uncertainties, a national databank

apparently legitimizing the whole technique "may be premature", says John Wadham, legal officer of the civil liberties group Justice.

Ironically, on the day before the commission's report was published, the court of appeal agreed to hear new evidence concerning two men, both of Afro-Caribbean origin, who were convicted of raping a number of students in Manchester in 1990 entirely on the basis of their identification through DNA samples.

In each case, defence lawyers argue that work since then by population geneticists suggests that the chances of a random match

between samples taken from the scene of the crime and those provided by the suspects were significantly underestimated. Franklin Sinclair, the solicitor representing one of those appealing against conviction, says that "until these uncertainties are sorted out, DNA profiling should not be accepted in the same way that we accept fingerprints."

But the general feeling is that continuing uncertainties in interpretation do not, in themselves, invalidate the case for a national database. The main priority is to ensure that adequate safeguards are provided for the way in which information is obtained, stored and used.

The royal commission itself has emphasized the need for such safeguards; and the police, who recently decided to destroy 2,000 DNA files because of uncertainty over their status, are keen to have a framework that will protect them from legal challenges.

**David Dickson**

## Challenge to growth hormone trial

**San Francisco.** A clinical trial in the United States of human growth hormone in normal children of short stature faces possible shut-down unless the National Institutes of Health (NIH) can show that it complies with federal regulations. A suit filed in late June in US District Court for the District of Columbia argues that the trial violates national rules for using children in clinical studies by needlessly risking patients' psychological and physical health.

The trial is controversial because the children are not deficient in growth hormone. Earlier studies have pointed to a possibility that these patients, after receiving additional hormone, may grow to less than their expected heights of up to 5-foot-6-inches as men or 5 feet as women. Critics also have raised questions about physical side-effects and the psychological impact of participating in the placebo arm of a study that could create distress about height.

The NIH suspended the trial last year to consider such questions, but resumed treatment in non-deficient children after two independent panels concluded that widespread use of the drug to treat prospective height-deficiency made growth hormone an important public health issue. They defended the experiment, in spite of what they considered greater than minimal risk for the children, on the grounds that unusual shortness is a pathological condition.

But the panels did warn against cosmetic use of growth hormone and said the study should stop making exceptions to include slightly taller young people than specified in the original protocol.

On 9 July, the Physicians Committee for Responsible Medicine, with 3,400 physician members, and a group led by anti-biotechnology activist Jeremy Rifkin, which filed the suit together, demanded of the

court documentation supporting the NIH view. The questions raised centre on the identification of short stature as a condition requiring treatment, NIH records of side-effects and the agency's relationship with the two manufacturers of the drug. NIH must respond within 45 days.

The NIH has no official comment on the suit, but Michaela Richardson, chief of the office of research reporting for the National Institute of Child Health and Human Development, said the US agency felt compelled to do the study because growth hormone, in spite of its high cost, is in use in some 7,500 non-deficient children and could possibly harm them. "We're not trying to cure shortness or anything like that," Richardson said. "We're just trying to understand whether this drug works so there can be a rational basis to prescribe it or not."

Interestingly, media coverage of the suit has prompted scores of calls from people wanting to join the trial. The NIH intended to enrol 80 short children but has been able to recruit only 36 by referrals from physicians.

The groups who filed the suit argue that the national agency is supporting a social stigma instead of attacking it, and carrying out the will of private drug manufacturers rather than meeting public need. Two companies have enjoyed a growing, and exclusive, market for human growth hormone, which sells for \$20,000 or more for a year's treatment. Genentech Inc. sold \$205 million of the treatment in 1992. Eli Lilly, which is providing drugs for the NIH trial, sold \$170 million in the same period.

Both versions of the drug are protected from competition until 1994, when they must find a way to retain market share. The treatment is being studied for other conditions, including osteoporosis and chronic renal insufficiency.

**Sally Lehrman**



# Japan explores the boundary between food and medicine

**Tokyo & London.** Japan's leading cosmetics manufacturer Shiseido is now marketing rice as a health product. This is the first step by Japanese industry to create a new market for foods engineered to have special medical benefits.

Last month, Shiseido became the first company in Japan to win approval from the Ministry of Health and Welfare to sell a "physiologically functional food", defined by new legislation introduced last September. Shiseido's product consists of rice from which the protein globulin has been removed for the benefit of those allergic to it.

For unexplained reasons, allergy to rice has become common in Japan, afflicting thousands of people young and old. The allergy causes unsightly red lesions on the skin covering large areas of the body. The present cure, the avoidance of rice and its products (including *sake*) in the diet is not welcomed by the Japanese.

Shiseido's engineered rice is one of many products being developed by hundreds of companies expecting to create a new niche in Japan's huge food market. Basic research in the field by university researchers is being supported by a large grant from the Ministry of Education, Science and Culture (MESC).

Shiseido's rice is produced with an enzyme that removes the allergen while retaining 80 per cent of the nutritious rice seed protein. Reconstituted rice is given a glossy surface; its developers claim that its taste is that of ordinary rice and that it prevents allergy in about 70 per cent of patients.

Another newly approved product is low-

phosphate milk, produced by Morinaga Milk Company for patients with chronic kidney disorders. Thirteen others are in the final stages of the eight-step approval process (see figure), including oligosaccharide-based foods for regulating intestinal flora, peptide-based foods for regulating mineral absorption and a material based on soya bean protein for regulating blood cholesterol.

According to Soichi Arai of Tokyo University's Department of Agricultural Chemistry, who is a member of one of the *ad hoc* committees set up by the Ministry of Health and Welfare, "at least 200 companies" are involved in the research and development of physiologically functional foods. Arai says that companies such as Nestlé are also involved. He is surprised that there seems to be less activity in the United States.

Although it is not illegal in Japan to sell products such as these as food, approval allows companies to claim medical benefits on their labels. Critics are worried that the approval process, which takes one to two years, will not be strict enough, but Shiseido's rice was tested on about 2,000 patients before approval.

The interest of the MESC is unusual. Arai heads a team of 57 researchers at 23 universities supported by approximately ¥700 million (\$6.5 million) over three years. One of Arai's studies centres on an inhibitor of cysteine proteolysis isolated from rice which can inhibit the proliferation of viruses such as herpes virus on skin and eyes; his group hopes to breed transgenic rice with high levels of the inhibitor.

**David Swinbanks & John O'Brien**

## Fish-oil miracle additive brings benefits to some

**Tokyo.** Japan's interest in health-improvement foods is illustrated by the current boom in sales of a fatty acid component of fish-oil, itself the product of people's contentment with their high IQ and their low incidence of heart disease, the curiosity of a fishmonger and the commercial enterprise of a researcher.

Kiyoshi Kondo, head of the Sagami Research Center southwest of Tokyo, explains that suggestions that "the Japanese might be so clever because they eat lots of fish" stimulated research on fish fatty acids and, in particular, on a material called docosahexaenoic acid (DHA), which was shown to reduce blood cholesterol levels and blood pressure and to improve the brain power of rats.

But the finding of a rich source of DHA was a happy accident. A local fish-shop owner asked Kazumagu Yazawa, a researcher at the centre, why his own mashed-up raw tuna was considered to be much more delicious than that of other shops. Analysis showed that the addition of fish eyeball fat from tuna was the explanation, and that these tissues are an unusually rich source of DHA.

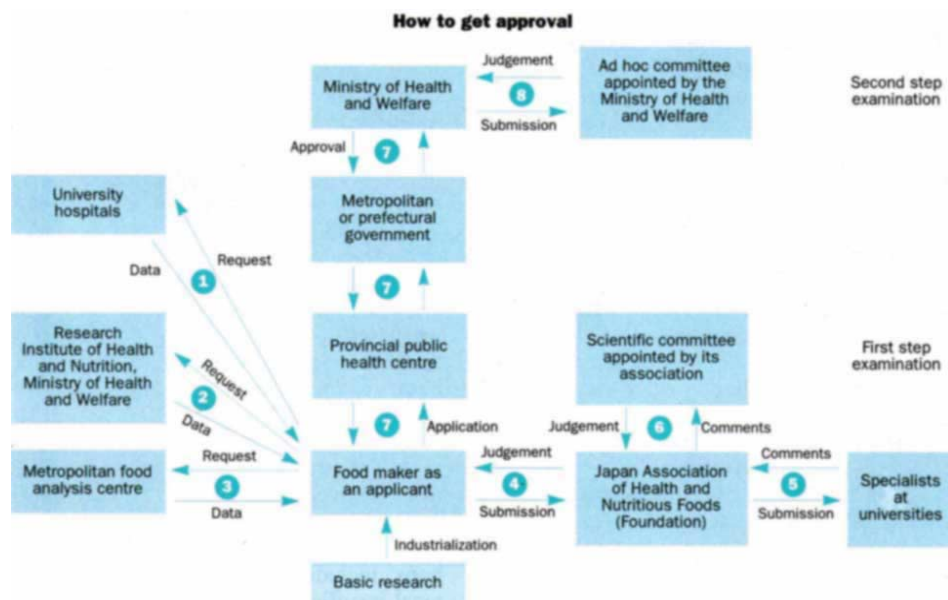
Indeed, the orbital fat of tuna and bonito contains 30–40 per cent of DHA, compared with the 5–10 per cent in other fish and can be easily purified. This has brought down the price of DHA extraction from several tens of thousands of yen for a gram to a few hundred yen (a few dollars).

Yazawa and his institute have linked up with a company to mass produce capsules of the fish oil, which they sell as a food supplement. The 8,000 bottles of capsules sold by the institute at ¥18,000 (US\$160) each represent only a tiny share of the whole DHA market, worth about ¥5 billion (\$45 million) a year.

Now the institute is also working with the small company Sano Shokuhin, which has been feeding the oil to chickens to make DHA-rich eggs. Farms all over Japan have followed suit. Yazawa predicts that many other DHA-enriched foods will soon appear at supermarkets and convenience stores. Moreover, clinical trials using DHA are also about to begin at the National Cancer Center in Tokyo because DHA has also shown antitumour activity.

Kondo says that an application for approval of the DHA-enriched foods as "physiologically functional food" is being considered, but booming sales hardly require the stimulus of approval.

**D.S.**





# American witnesses testify in Japan about AIDS risks

**Tokyo.** Litigation by Japanese haemophiliacs against the Japanese government and blood product manufacturers reached a crucial stage last week. The patients are seeking compensation for failure to protect them from blood products contaminated with human immunodeficiency virus (HIV).

The latest development is the testimony in Osaka district court by a former official of the Centers for Disease Control (CDC) at Atlanta in the United States. Donald Francis, former assistant director of viral diseases and coordinator of the AIDS laboratory at CDC, testified that by early 1983 it was clear that US blood supplies were probably contaminated with the transmissible agent responsible for AIDS.

Yet, during 1983–85, Japan dramatically increased its imports of untreated US blood products (see figure) and, as a result, many of Japan's 5,000 haemophiliacs were infected with HIV. Francis said the tragedy could have been avoided.

As in France (see *Nature* 363, 491; 1993), Japan was slow to react to the dangers posed to blood supplies by AIDS. Blood coagulant concentrates for haemophiliacs which had been heat-treated to kill HIV and other viruses were not approved for sale in Japan until July 1985, despite their availability in the United States from March 1983. Blood product manufacturers continued to advertise untreated products in Japan without warning until late 1985; Japanese haemophiliacs used them into 1986.

The Osaka suit was filed in 1989 by a group of haemophiliacs and seeks about \$1 million each in compensation from the government and five pharmaceutical companies — Green Cross Corporation (Midori Juji), the Chemo-Sero-Therapeutic Research Institute (Kaketsuken), Baxter Travenol, Bayer Yakuin and Nippon Zoki Pharmaceutical Company. Together with a similar suit filed slightly later in Tokyo, the number of plaintiffs has reached 92, but several have died and are now represented by their families.

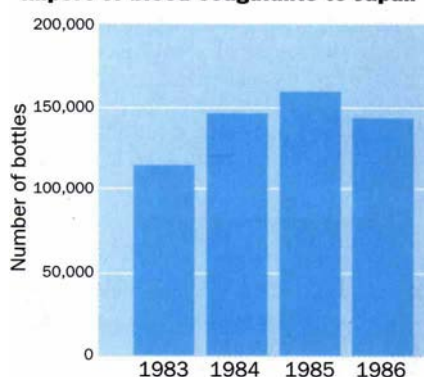
Francis says of his court experience that he was "almost tearful" in the presence of the infected haemophiliacs. He says that by early 1983 the danger posed to US blood supplies by AIDS was "absolutely black and white", when six cases of AIDS in five adults and one baby had been traced to transfusions of blood from donors who subsequently developed AIDS.

Francis said that CDC convened a meeting in January 1983 to explain the dangers to blood banks, blood product manufacturers and physicians, and to suggest possible solutions. "But we were on totally different wavelengths." The blood industry argued that six cases of AIDS in 6 million donors

indicated a one-in-a-million chance of infection. Francis recalls that the meeting ended with his asking, "How many more people have to die before you understand?"

Despite these warnings, blood product manufacturers increased their imports to Japan of untreated concentrates as Japan began a massive "modernization" drive to introduce concentrates for use by haemophiliacs in the home. The price of the concentrates was then much higher in Japan than in the United States, providing a strong incentive to sell in Japan, especially after the United States began to withdraw untreated

Import of blood coagulants to Japan



Source: Mainichi Newspaper (12 March 1988)

products in 1984. Doctors and hospitals in Japan also profited by buying products at discount and claiming the higher official price from insurance schemes.

According to Francis, lawyers for the plaintiffs have shown him a memorandum from one importer saying that the company had consulted CDC and had been assured there was no danger to Japanese blood product supplies. "I don't know who they spoke to but it must have been someone in the lead poisoning department."

"Inability to translate hot cutting edge news in science to practical health care programs" also played a key role in the tragedy, Francis says. "There were only a dozen experts on AIDS in the world at that time.... It was in part a failure in marketing." This view seems to be borne out by Shelby Dietrich, head of the Hemophilia Center of Huntington Hospital in California and former chairman of the medical board of the World Federation of Hemophilia, who has testified for the defence in Tokyo. She said she was not convinced that a blood-borne virus was the primary cause of AIDS until a paper was published in *Science* by Robert Gallo and colleagues in May 1984. And the first major change in treatment of haemophiliacs was not made at her hospital until September

1984, when it became clear that heat treatment killed HIV.

Dietrich's testimony reveals the problems that doctors faced in weighing up the risks and benefits of various blood products in 1983–84. After the January 1983 CDC meeting, her hospital reverted to treatment of limited groups of patients (the young and those with mild haemophilia) with cryoprecipitate made from a much smaller pool of donors than concentrates. But the cryoprecipitate requires treatment in hospital. Complications due to delay in treatment can occur and one patient died as a result. Many patients then chose to revert to the concentrates which can be self-administered.

When heat-treated concentrates became available in the United States in March 1983, Dietrich and other doctors initially used them only in a limited number of patients because of fear of side-effects. Only when heat treatment was shown to kill HIV in September 1984 did their use become widespread in the United States.

In 1983, the risk of AIDS infection from blood products appeared to be small. People failed to recognize the dangers of the long incubation period of at least two years established by the evidence from the 1982 transfusion cases, Francis says. He blames the US Food and Drug Administration (FDA) for failing to act quickly enough. "We at CDC tried to pass the baton to FDA at the January 1983 meeting", but "Reagan was decreasing budgets and the government turned the responsibility over to the blood banks". The "terrifying thing", he says, "is that this worldwide disaster is the result of decisions made by a remarkably small group of people led by Americans."

At the 1983 meeting, Francis says he advocated the introduction of screening of blood donations for hepatitis B as an anonymous way to eliminate most male homosexual donors, a high-risk group for AIDS, from the donor pool. But this idea was not adopted. And effective screening of US blood supplies did not come until the introduction of the HIV antibody test in 1985.

Alarmed by the reports from the United States, Japan's Ministry of Health and Welfare set up a committee of academics in 1983 to look into the possibility of reintroducing cryoprecipitate instead of concentrates. In March 1984 the committee opted for concentrates. Clinical trials of heat-treated concentrates began around this time in Japan but were not approved for use until more than a year later.

Ministry officials argue that they acted as quickly as possible. But Francis and lawyers for the Japanese plaintiffs point out that the dangers of viral contamination of concentrates made from the pooled blood of thousands of donors were foreseen in the 1970s before AIDS even appeared, and a German company developed heat treatment in 1978 to eliminate hepatitis infection. Court cases like this in Japan typically take years to resolve.

David Swinbanks

# Food technology divides British teachers

**London.** If Britain's technology is eventually regenerated, historians may place the beginning of the trend in 1993, and identify its vehicle as the soggy flapjacks (cereal bars) now being produced in British schools as part of the new technology curriculum.

For some months, 15-year-olds throughout England and Wales, male and female alike, have been struggling to devise a recipe for a "high energy snack", one of the required tasks in the technology curriculum introduced three years ago. But the question of whether this is a wise use of students' and teachers' time has helped to trigger a conflict between teachers and the government just as bitter as recent rows over the compulsory assessment of basic skills.

The problem started in 1988, when the government decided that all children should be required to spend part of their school career studying technology. Home economics teachers, seeing the prospect of their subject being marginalized as a result, fought — and won — a battle to ensure that cooking and nutrition, under the label of "food technology", should be included in the new

compulsory curriculum, introduced in 1990.

Three years on, the teaching of technology in schools is in disarray. Last year, faced with complaints from both teachers and potential employers of an unwieldy and unfocused syllabus, the government promised to introduce a new revised curriculum from the autumn of 1994 for children up to the age of 13, and in the following year for the 14-to-16 age group.

Last week, however, faced with continuing disagreement over what and how much this curriculum should contain, John Patten, the Secretary of Education, announced that the introduction of new technology curricula would be delayed for another year in each case, to provide time for further revision. Students will still take the heavily criticized existing courses next year, but will not now be assessed on them.

At the heart of the dispute is the lack of a sharp definition of what "technology" meant when the national curriculum was established. Last December, the Department for Education published a set of proposed changes that attempts to provide a tighter

focus for the curriculum, and greater specification of its essential teaching content.

The National Curriculum Council (NCC), in an interim response to these proposals two weeks ago, described them as a "marked improvement" on the current curriculum, but said they did not go far enough. In doing so, the NCC was responding to criticism from, among others, potential employers. This was expressed through organizations such as the Engineering Council, which has argued that technology courses should be strongly structured around manufacturing skills and industrial production.

Some teachers endorse this approach. But others complain that it is too narrow, and risks alienating precisely those who would benefit from increased familiarity with a range of technological processes, not merely those that belong to the factory or the office. "We argue that technology should be defined in the broadest possible way", says Sheila Denton, assistant secretary of the Association of Teachers and Lecturers.

The dilemma is epitomized by the issue of food technology. The Engineering Council argues that while certain aspects of food production "lie properly within technology", cookery and nutrition should be rescheduled as life skills. Home economics teachers, however, remain adamant that their subject has as important a role in teaching students to handle "soft materials" as engineering does in handling "hard materials" such as wood and metal — and thus legitimately belongs to the technology curriculum. "We believe that food represents a user-friendly approach to the whole field of technological activities, using materials that are familiar to everyone", says Geoffrey Thompson, secretary of the National Association of Teachers of Home Economics.

Much will now depend on the direction in which Sir Ron Dearing, the chairman of the Schools Curriculum and Assessment Authority set up by the government to inquire into the future of the national curriculum, decides that the technology component should be steered. At present, the odds are on the Engineering Council's approach, which is supported by many school science teachers, concerned at what they see as a lack of theoretical underpinning to the current technology curriculum.

A final response from the NCC is due in the autumn, and the council says that the question of whether food should be included is still undecided. Meanwhile, the council has already agreed to provide schools with examples of technology projects "in a wider range of materials" than those currently being used. That could mean the return of traditional skills, such as woodwork and metal-working, and the consignment of soggy flapjacks to the education history books.

**David Dickson**

## Recession causes fall in R&D spending

**London.** Figures released last week by the Organization for Economic Cooperation and Development (OECD) confirm that the economic recession has caused a significant

slowing down in the research and development (R&D) expenditures of all advanced industrialized nations.

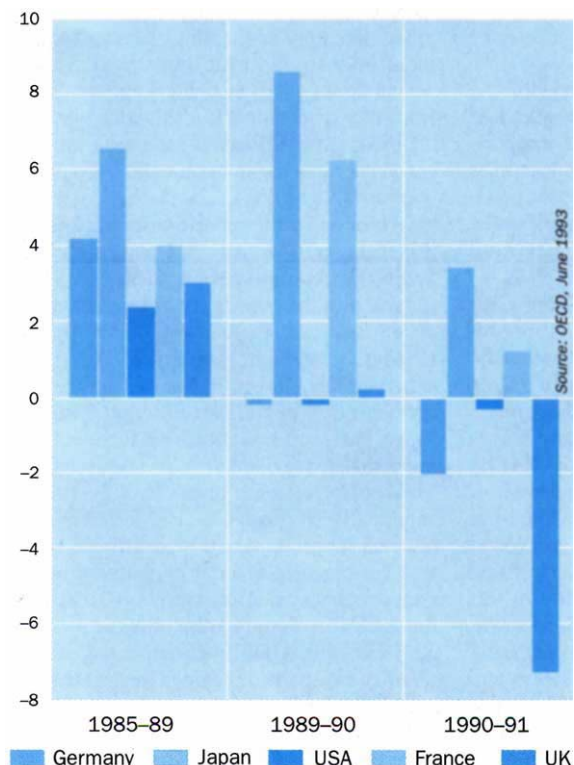
Spending on R&D in the United States, measured in prices adjusted for inflation, fell by 0.3 per cent between 1990 and 1991. In Japan, R&D continued to grow by 3.4 per cent in 1990–91; but this was in sharp contrast to an 8.6 per cent increase in the previous year, and it also resulted in total spending on R&D dropping from 2.88 to 2.86 per cent of gross national product (GNP).

In Europe, the biggest reductions in 1990–91 was in Germany, whose spending on industrial R&D dropped by 2 per cent, and in Britain, where a fall of 7 per cent is blamed partly on cutbacks in defence R&D. France, in comparison, increased its R&D spending by 1.2 per cent; but the total still fell from 2.41 to 2.40 per cent of GNP.

Among the middle-sized OECD countries, only Denmark and Austria achieved significant increases. Spending on R&D in most of these remained roughly constant, although it fell by 6 per cent in Sweden.

**D.D.**

**Average annual % growth in gross domestic expenditure on R & D at constant prices**





# Reform options for peer review

SIR — Ernst, Saradeth and Resch (*Nature* 363, 296; 1993) present valuable but worrying data on the reliability of the peer-review system of a medical journal. If their findings were found to be the rule rather than the exception, a number of reforms merit consideration: (1) there is a strong case, as suggested by the authors, for all reputable scientific journals to conduct annual blind tests of the quality of their reviewers; (2) the number of reviewers could be increased with the aim of increasing reliability; (3) reviewers could be identified to authors in an attempt to increase accountability and encourage open debate; (4) there seems to be no reason for submitted manuscripts to indicate the name or institution of the authors.

In addition we need to openly address the question of why these inconsistencies occur. Finally, in agreement with the authors, we emphasize that a balanced review system is essential for scientists, whose careers are strongly influenced by acceptance or rejection of publications.

**R. H. Bradshaw**

**N. E. Bubler**

*Animal Behaviour Research Group,  
Department of Zoology,  
South Parks Road, Oxford OX1 3PS, UK*

SIR — Ernst *et al.* suggest drawbacks with the peer review of academic articles, but overlook authors while drawing attention to a matter of concern for editors. What they describe as reproducibility is the persistent problem of reliability between markers of any examination. They describe a criterion referenced marking system and deduce, quite reasonably, that perfect reliability is seldom if ever achieved.

Any teacher or student in higher education will know that there is always going to be a variance between the judgements of individuals where there is a high degree of inference. The criteria Ernst *et al.* suggest, vary considerably in the degree of inference required; however, the criteria they suggest are all high inference variables. In practice, high reliability is associated with low inference. Low inference is often used with matters of little importance.

Take for example "linguistic merit" — one would not reasonably expect there to be a high measure of reliability between expert raters unless there were ample exemplars of the anchor points to the scale. Examples that instanced "high linguistic ability" would need to be contrasted with instances of "low linguistic merit" — and the other points on the scale would require similar exemplars. With criterion referenced assessments, reliability can be improved with training and models.

Your correspondents miss a significant point with regard to editorial responsibility.

Editors of learned journals exercise judgement in order to maintain standards in science. They do so imperfectly, but they accept, along with most readers, that the final judgement must lie with the reader. Errors are likely to occur in any activity mediated by humans. Examinations are no less reliable or valid than refereeing scientific publications. The lesson to learn is that reliability can be improved if models of performance are made explicit and those making judgements are properly trained.

**Ray McAleese**

*Institute for Computer Based Learning,  
Heriot-Watt University,  
Edinburgh EH14 4AS, UK*

SIR — Ernst *et al.* are a little subjective in their analysis of the results from their anonymous peer-reviewing exercise. They sent the same paper to 45 experts who were asked to assess the manuscript according to nine quality criteria. Ernst *et al.* tell us that the reviewers' responses were so poorly reproducible as to render peer review unacceptable.

So let us review their study. Out of 45 potential referees, only 31 (69 per cent) returned adequately completed reviews. Is this response rate unacceptable, acceptable, fair or good? It can scarcely be excellent. Further, although the data are rather poorly presented and described (quality of table = unacceptable), it seems that in only two of the nine criteria did all 31 respondents offer an opinion (scientific merit and statistical methods — which says something about current vanities). In six of the criteria, at least one of the respondents failed to offer an opinion. Two (6.5 per cent) of the respondents were even unable to make a comment about their overall judgement of the paper. These inadequacies are not discussed by Ernst and colleagues. I would thus judge the methodology and statistical methods of their study to lie somewhere between acceptable and unacceptable.

And what of the discussion? Ernst *et al.* write: "the results . . . demonstrate disappointingly poor reproducibility with extreme judgements ranging from unacceptable to excellent for most criteria". But I see it differently. On the basis of the results given, only 4.4 per cent of the opinions suggested the paper was unacceptable; 21.0 per cent of the opinions suggested the manuscript was excellent and 74.6 per cent found it acceptable, fair or good. In other words, three-quarters of the referee sample thought the paper was worth publishing, with modification. This is not an extreme range of opinions but pretty good agreement.

Some readers might think it a mite unfair to submit a short letter to *Nature* to

this type of criticism. But it was submitted as a serious study of the peer-review system in order to invite comment, raise awareness, stimulate further research and so on. This then raises the question of whether *Nature* should have published such a questionable study. And the answer, of course, is "yes". The study may have flaws but it is interesting. Indeed, although the sharply critical among us might be tempted to criticize the study into the ground and deem it unacceptable, more generously minded individuals might say that the conclusions and methodology are acceptable or even good and some (perhaps those who have a grudge against the peer-review system) might hail the study as excellent. For, like it or not, the peer-review system is little more than a market research exercise (using a very small sample) to judge the response of the entire population (scientists likely to read the paper) to the matter in hand. Regarded as a crude opinion-polling exercise, peer-reviewing can be easily improved. Editors need to base their publishing judgements on the opinions of quite a few reviewers (probably five or six — but don't ask for statistical justification). They also need to make sure that they constantly vary their reviewers so that the sample remains representative. (Why should one or two individuals control all the papers on a particular subject published in a journal?). And editors need to ask their reviewers the right questions ("Do you agree with the authors' conclusions?" rather than "Comment on the scientific merit"). By contrast, regular, blind tests of the quality of reviewing, as Ernst *et al.* demand, may be well-intentioned but are futile.

**Simon P. Wolff**

*University College London Medical School,  
Rayne Institute,  
5 University Street,  
London WC1E 6JJ, UK*

SIR — Ernst *et al.* describe the poor reproducibility of the peer-review system for publication or rejection of submitted articles. For my part, I would take encouragement from these statistics, which suggest that the selection of referees has little impact on the outcome.

Of greater concern is the finding that the main reason for rejection of the paper was criticism of the statistical methods employed; four referees assessed the statistical methods to be unacceptable, but seven judged them to be excellent. Application of a statistical method to analyse a particular dataset is either right or wrong and the 20 referees who thought that the statistical methods were acceptable, fair or good were probably not sure.

There are cases in peer-reviewed journals not only of the application of the wrong statistical methods but also of misinterpretation of the statistical findings. I



am also aware of articles with correct statistical methods that have been rejected because, in the referees' opinion, the statistical methods were not appropriate. Submitting a manuscript again, together with a letter explaining and justifying the statistical methods, is not acceptable to authors. The data of Ernst *et al.* reinforce the need for reputable learned journals to refer all statistical methods (including interpretation of statistical analyses) to a reputable statistician.

**P. Gale**

Water Research Centre,  
WRc Medmenham,  
PO Box 16 Marlow,  
Buckinghamshire, SL7 2HD, UK

SIR — The system that Ernst *et al.* tested is flawed from the start. To ask reviewers to judge a submitted paper merely numerically is to invite the disparity these authors found. Worse, it provides the editor with no basis for making an independent judgement (except for calculating an average), and provides the author with no basis for improving the paper. Editors of the journals published by the Entomological Society of America (whose *Annals* I have co-edited for 20 years) require reviewers to write — in words — what they think of a paper, and why. The editor can then evaluate both the reviews and the paper; and the author then knows what needs to be improved, or why the paper will not be published. We editors have found this system useful and reliable, and authors have found it beneficial.

The peer-review system has been challenged by many people over many years. Yet no-one has devised a better method for judging submitted papers. The peer-review system is rather like democracy, about which one Anglo-American remarked that it is the worst of political systems — except for all the others.

**Carl W. Schaefer**

Department of Ecology  
and Evolutionary Biology,  
University of Connecticut,  
Storrs, Connecticut 06269-3043, USA

SIR — The aims of Ernst *et al.* may be noble (to improve the peer-review system) but their conclusion (that the system is unreliable and lacking in reproducibility) cannot be justified by their mini-experiment. The data can have another interpretation.

The bottom line for many academics who submit manuscripts for publication is whether they are accepted. Whether a paper is considered acceptable, fair, good or excellent by the referees is irrelevant; that will be decided over time by the academic community as a whole.

Once these categories are combined into a single 'acceptable for publication', the data from Ernst *et al.* show that 29 out

of the 31 referees thought the paper acceptable while two did not, implying 94 per cent agreement and 6 per cent dissension. Surely that indicates clearly that the peer-review system does work? It certainly does not indicate that the system is unreliable. The consideration of most manuscripts submitted to peer-reviewed journals by more than one referee only strengthens this argument.

**J. Richard M. Thacker**

Biology Department,  
University of Paisley,  
High Street, Paisley PA1 2BE, UK

SIR — Ernst *et al.* express concern about the reproducibility of the peer-review system. On the basis of their data, I conclude that peer review is actually quite reliable. Most journals routinely solicit two reviews for each manuscript. When two reviewers disagree substantially on the merit of a manuscript, the paper is often sent to a third reviewer. In the test case presented by Ernst *et al.*, choosing any two reviews randomly from the 31 available would probably yield two generally favourable, reasonably consistent reviews. If an unfavourable review was chosen initially, the probability that a third review, again chosen at random from this group, would be favourable is quite high. The probability that this manuscript would have been viewed very unfavourably (two overall judgement scores of 1 or 2) based on two randomly selected reviewers from this group is obviously not zero; it is 0.017. Do we expect better reproducibility than this from any living system?

Of course, there is always an element of subjectivity in any human judgement, and so there is a chance that a scientifically sound manuscript will be rejected inappropriately under the existing peer-review system. The exercise undertaken by Ernst *et al.* suggests that the chance is surprisingly and reassuringly small. For all its human failings, peer review seems to be alive and quite well.

**Samuel A. Green**

University of California, San Francisco,  
Hormone Research Institute,  
San Francisco, California 94143-0534, USA

SIR — Ernst *et al.* show clearly that, even when a paper is of high quality, getting it accepted is like driving on a bumpy road: one may finish in the ditch. If we want a better system of evaluation, we must replace the present lottery.

We think that an open system would help. A few lines at the end of the paper, saying who accepted the paper (as is often done already), who reviewed it and the main questions raised, would benefit the referee, the author and the journal, as well as inform the reader. Referee work would thus gain kudos and readers could compare their opinions with those of the authors and referees, giving the reading of

a paper an additional attraction: the possibility of judging the referees.

Indeed, a few journals already ask authors to suggest potential referees or publish author/referee discussions of invited papers, thus going in the direction we are suggesting.

**Maria Gabriella Manfredi Romanini**

**Carlo Alberto Redi**

University of Pavia,  
Department of Animal Biology,  
Piazza Botta, 10, 27100 Pavia, Italy

SIR — The communication by Ernst *et al.* proves that the peer-review system is gravely flawed. But I have the solution. In January 1984, I submitted a proposal to the US National Science Foundation (NSF) requesting funds; five months later, the foundation sent me my own proposal for review. I naturally jumped at the opportunity to break new ground and sent back a totally objective review. In fact, my review was so objective that the proposal was rejected.

Since then, I have been conducting an in-depth study of the peer-review problem, a problem so vexing that it has even attracted the congressional eye (Stone, R. *Science* **256**, 959; 1992). This study has led me to conclude that not only proposals but also articles and books should be reviewed by their own authors. Recognizing that the author is of course the person most qualified to understand what he (or she) is doing, has done or plans to do, I hereby propose the adoption of the Absolute Review System (ARS), a system in which the authors themselves (the Absolute Reviewers) do the reviewing.

The advantages are several. First, the material will be reviewed, for a change, by a person really qualified to do the reviewing. Second, there will be no spread of opinions, instantly eliminating the problem that so pains Ernst *et al.* Third, the state of total nervous collapse in which programme directors and journal editors find themselves each time they try to get reviews back in time will simply vanish, because the authors will be jumping at the opportunity to review their own work. And fourth, peers everywhere will be freed from the demoralizing task of having to review works they do not understand but which they review anyway just to show that they actually understand them.

My review of my own proposal has been published in the *Journal of Irreproducible Results* (**37** (6), 12; 1992) and my review of my latest book has been published in *EOS* (**74**, 258; 1993). Anxious authors, journal editors, programme directors and peers are referred to these publications for guidance, solace and a glimpse at the freedom that the future promises.

**Cesare Emiliani**

Department of Geological Sciences,  
University of Miami,  
Coral Gables, Florida 33124, USA

# The kinship of apes and people

**An unpromising argument that apes should forthwith be regarded as equal members with people of a single community deserves a wide hearing, not because it is likely to happen but because it is constructively provoking.**

To borrow from the bookmakers, what follows is a turn-up for the book. Not often is it possible to pick up a volume which, on the strength of its preface and blurb, promises to be an unmitigated irritant, and then to find that the text is compulsive reading even if the argument falls short of being compelling. Yet that is the case with *The Great Ape Project*, edited by Paola Cavalieri and Peter Singer (Fourth Estate, London, 1993). Singer is the better known of the two editors, and he chiefly as the one who has made the term *speciesism*, which stands in relation to animal species as does racism to the human race, one of the most common terms of abuse in the vocabulary of the animal rights movement. (The inventor of the term is one Richard D. Ryder, described as a British lobbyist in the Liberal Democrat cause, who is one of the 34 contributors to this symposium volume — and the briefest.)

The reason why the book threatens to be an irritant is that it begins with a tract, or manifesto, with the sententious title "A declaration on great apes". (The reasons for the lack of a definite article before "great apes" will become clear.) The manifesto says that the time has come to admit into the "community of equals" not only members of the human species but orang-utans (*Pongo pygmaeus*), gorillas (*Gorilla gorilla*) and chimpanzees of both known species (*Pan troglodytes* and *P. paniscus*). And the community of equals? That is the set of living creatures belonging to these species to all of whom the respect (in the legal sense) at present accorded to members of the species *Homo sapiens* is to be extended.

Varying the terms of the Declaration of Independence only slightly, the authors specify that the moral principle of the community of equals should be life, liberty and the "prohibition of torture". A few caveats quickly arise. Great apes, human or otherwise, should not be killed "except..., for example, in self defence". To deprive members of the community of equals of liberty is not allowed except for their own protection or that of others, in which case there must be an appeals procedure, if necessary by advocates as in the present legal treatment of members of the human community. And "no torture" means "no experiments without informed consent", which only adult members of the relatively articulate *Homo sapiens* can provide.

Just how fanciful is the notion of the community of equals? The two editors (who add that all their contributors "subscribe" to their declaration), acknowledge that not all

members of *Homo sapiens* now benefit from their moral principles, so that there may be little enthusiasm for extending them to other species. But on laboratory experiments with non-human apes, they insist that the balance of the argument has so shifted that the onus of proof lies with the experimenters. (Neither "vivisectionist" nor its variant "vivisector" is conspicuous in the text.) To guard their rear against the likes of those who are forever "liberating" beagle dogs from laboratories, they offer the community of equals as a possible first step towards the inclusion of other species. (But that has not convinced Steve F. Sapontzis of the California State University, who protests that singling out the non-human apes for special treatment simply perpetuates the errors of speciesism.)

This is an arid starting-point for a good read. There are two reasons why the book succeeds in spite of its agenda. One is that most of the contributors have taken their assignments seriously, raising a string of interesting questions in the process. Another is that the bulk of the book necessarily encircles a question that becomes more challenging with each year that passes: what precisely are the behavioural differences between people and other apes (to follow the Cavalieri/Singer idiom for much of what follows), what features of the genome do they represent and how did they arise genetically?

That said, most of the contributors enter fully into the spirit of the agenda set for them, which is to demonstrate the kinship between people and apes. There is plenty of material. People who have spent time with chimpanzees, such as Jane Goodall and Toshisada Nishida, tell what that is like with humour.

Those who converse with apes (by sign-language, not natural for either party) and who judge their emotional states by looking into their eyes occasionally protest too much that they are aware of the pitfalls of anthropomorphism, but that does not crab their tales. That apes (not just chimpanzees), or at least some of them, can communicate with people must be taken as a fact. That they can do so with a measure of abstraction, distinguishing between the first, second and third persons, is something to brood upon.

Richard Dawkins has the most elegant resolution of that question: the sharp discontinuities attributed by taxonomists to the gulfs that separate species are, in the sweep of evolution, an illusion.

As one goes back along an evolutionary tree, or a cladogram, the distinctions between extant species are blurred, and who can tell with confidence which existing popu-

lation of a supposedly single species harbours the founders of further speciation?

Even so, Dawkins is not the only contributor to marvel at the shortness of the timespan occupied by the evolution of the apes; 15 million years lumps us all (in the Singer idiom) together, an extra 5 million years would include the gibbons as well. And everybody marvels at the homology of cognate genes among the apes, whence Jared Diamond's name "the third chimpanzee" for people and his question how visitors to zoos will react when the chimpanzees are housed in cages labelled *Homo troglodytes*.

But, curiously, nobody makes much of the circumstance that people differ from other apes in having only 46 chromosomes, not 48. That is surely relevant to the agenda of the community of equals in two practical ways: it is a reminder that not only the structure of the genes matters, but also their arrangement and, presumably, coordination. And it suggests that genome sequencing may also eventually explain why some of the phenotypic differences between different apes (Singer's idiom) are sharper than homology would suggest.

Meanwhile, readers must be content with what philosophers, sociologists and lawyers have to say. Indeed, they turn out to be more interesting than the biologists, no doubt because they cannot dodge the management of the community of equals. Their consensus is that it is in principle like the management of human society, in which the disadvantaged deserve respect and compassion but in which there are also many more of them, the members of what we persist in calling different species.

In the Latin sense of the word, the community of equals is an admirable cause. It is also an interesting and provoking light in which to put relationships with other animals. One of the questions that arise is whether, in a future community of that kind, the smarter set would compassionately embark on improving the genetic lot of the disadvantaged, and what the consequences would be. Most of the contributors to this symposium would probably be offended by the thought, but when the techniques are safely available, it is not easy to see why they should be. Much the same dilemma arises already in the question whether isolated tribes in the Amazon should be helped to a more modern way of life; the tribesmen are usually more eager for change than those who might help them. That is what undermines the community of equals we like to think we have already.

John Maddox



# Don't touch that dial

J. W. C. White

AT a time when superlatives are routinely used to describe the mundane, it is difficult to express the importance of two papers in this issue which present results from the new GRIP ice core in central Greenland. The GRIP Project Members<sup>1</sup> and Dansgaard *et al.*<sup>2</sup> (pages 203 and 218, respectively) give us our first detailed look at the last interglacial period, and it is not what we expected.

As an uncertain climate lies before us, we have been looking warily over our shoulders to see how the climate system has behaved in the past. For 10,000 years, the Earth has enjoyed an interglacial period, a time of steady and dependable climate. Further back, during the last ice age (which lasted about 100,000 years), and in the transitional period, it is now accepted that the climate 'flickered' rapidly. But we could take comfort from the thought that dramatic changes occurring in decades or even years were probably triggered in some way by the massive glaciers or huge extensions of sea ice present at the time.

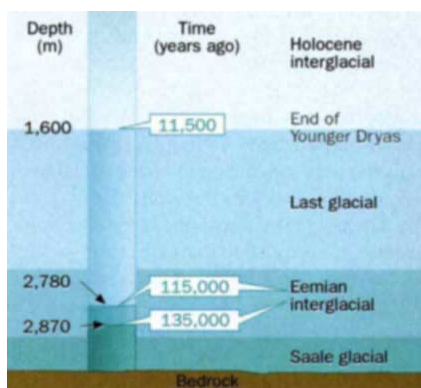
Now the Greenland Ice-core Project (GRIP) team have removed this sense of security. Using a variety of evidence drawn from the ice core — stable isotope ratios, chemical and physical properties, and greenhouse gas concentrations in trapped air bubbles — they demonstrate that very rapid shifts in temperature and greenhouse gases are also possible in interglacial periods.

Our view of climate on the timescale of glacial cycles is shaped by the tools we use to construct that view. Up to now, our ideas of interglacial periods have come from three sources. First and foremost is our knowledge of our own Holocene interglacial, drawn from sources such as tree rings, historical records and pollen samples, to name but a few. Then there are the ocean sediment cores which document numerous interglacials over the past million years. Third, there is the Vostok ice core, until now the only deep core to have yielded easily datable ice from the previous interglacial period (known as the Eemian) which stretched from about 135,000 to 115,000 years ago.

In ocean cores the stirring of the sediment surface by benthic life homogenizes the oceanic record so that the minimum resolution is no better than a thousand years. In the Vostok core, the low accumulation rate of snow and thinning of the annual layers with depth means that climate changes of a century or less are difficult to resolve; flickers may have occurred that cannot now be detected. (There is a compensation: the Vostok core should take the record back to about

500,000 years ago, twice the age of the oldest Greenland ice<sup>3</sup>, and covering several glacial to interglacial cycles.)

The new Greenland ice cores GRIP and GISP2, in contrast, were drilled in regions of high snow accumulation near the centre of the Greenland ice sheet. With these cores, designed to concentrate on the past 200,000 years, we can see climate changes on the timescale of decades or less, even though they occurred a hundred thousand years ago. The indicators of climate change range from local (for example, temperature, deduced from its effect on stable isotope ratios), to regional and



Depths of time — the GRIP ice core.

hemispheric (such as airborne dust concentrations and chemical composition), to global (greenhouse gas compositions).

The Eemian period falls in the interval from 2,780 to 2,870 cm down, well above bedrock. Blurred in the Vostok core, it now comes into focus and it is strikingly different from the Holocene. Holocene climate appears to have one, and only one, state, whereas the new results show that the Eemian had three. The middle state matches our own Holocene climate. A significantly colder state and a significantly warmer state existed in the Eemian. On average, temperatures were 2 °C higher than at present. It apparently took very little time, perhaps less than a decade or two, to shift between the states, and the states appear to be stable sometimes for thousands of years and sometimes for only decades. We don't know which is the norm for interglacial periods: the stable, one-state Holocene or the multiple-state, rapidly changing Eemian. We do know that answering this question will be a priority for global change research.

When evidence from the sister core of GRIP, the GISP2 core<sup>4,5</sup>, showed earlier this year that aspects of the climate system could shift from glacial conditions to interglacial conditions in a few years, there was always the solace that such changes were

characteristic of glacial times, and not really analogues of the future. In our interglacial age, we do not expect the polar front in the North Atlantic rapidly to dip down to Spain with sea ice expanding in behind it, plunging adjacent land, particularly Northern Europe, into glacial-like cold. We do not have massive lakes formed by retreating glaciers, lakes which may catastrophically drain into the North Atlantic, disrupting deep water formation and the transfer of heat northward.

The new ice core results bring rapid climate change to our doorstep: changes of up to 10 °C in a couple of decades, or perhaps in less than a decade, appear possible in interglacials. Given our ongoing 'global experiment' of increasing greenhouse gas concentrations via fossil fuel burning, is the Eemian warm state a glimpse at our future climate? Whatever the answer to that question, the speed with which the climate system can shift states gives us pause. Adaptation — the peaceful shifting of food growing areas, coastal populations and so on — seemed possible, if difficult, when abrupt change meant a few degrees in a century. It now seems a much more formidable task, requiring global cooperation with swift recognition and response.

How unusual is the climate stability of the Holocene? Dansgaard and colleagues<sup>2</sup> investigated one of the possible tracers, the oxygen isotope ratio (a proxy for atmospheric temperature) along the whole of the GRIP core (see figure). Throughout the last glacial period, the Eemian interglacial and the glacial before that, they found rapid oscillations in the isotope ratio. Because of the way that ice thins with age, they could look in detail at the Holocene, and found that the swings have been much smaller, by a factor of 3 to 4, than those earlier flickers. At no time during the Holocene has Eemian-like climate change occurred.

We humans have built a remarkable socio-economic system during perhaps the only time when it could be built, when climate was stable enough to let us develop the agricultural infrastructure required to maintain an advanced society. We don't know why we have been so blessed, but even without human intervention, the climate system is capable of stunning variability. If the Earth had an operating manual, the chapter on climate might begin with a caveat that the system has been adjusted at the factory for optimum comfort, so don't touch the dials. □

J. W. C. White is at the Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado 80309, USA.

1. GRIP Project Members *Nature* **364**, 203–207 (1993).
2. Dansgaard, W. *et al.* *Nature* **364**, 218–220 (1993).
3. Jouzel, J. *et al.* *Nature* (in the press).
4. Alley, R. *et al.* *Nature* **362**, 527–529 (1993).
5. Taylor, K. C. *et al.* *Nature* **361**, 432–436 (1993).



# Twins and T-cell responses

Brian L. Kotzin

A TRADITIONAL way of studying the interplay between genes and environment in the development of autoimmune disease takes a thoroughly modern turn on page 243 of this issue<sup>1</sup>. There, Utz and colleagues report a study of T-cell responses in identical twins in which neither, both or one of the twins had the neurological disease multiple sclerosis (MS). They find that in this third case the twin without MS has a different T-cell receptor repertoire in response to specific antigens compared to that seen in the twin with the disease. Here then may be more evidence of the contribution of non-genetic factors to autoimmune disease, but interpretation of the results is far from clear cut.

T cells are thought to be pivotal in the development of several autoimmune disorders, including rheumatoid arthritis and type 1 diabetes mellitus, as well as MS. Both genetic and environmental factors have been implicated in the disease process. In MS, elements of the genetic predisposition include particular class II alleles of the major histocompatibility complex, and (less certainly) polymorphisms within the T-cell receptor (TCR)  $\alpha$ - and  $\beta$ -chain gene complexes<sup>2,3</sup>. The nature of the association between TCR genes and disease remains controversial, but it seems that CD4<sup>+</sup> T cells bearing  $\alpha/\beta$  TCR are somehow involved. Further evidence on that score has come from an animal model in which experimental autoimmune encephalomyelitis (EAE) can be induced. Immunization of animals with proteins of the central nervous system, such as myelin basic protein (MBP) or proteolipid protein<sup>4</sup>, has effects which in some respects resemble those of MS in humans. CD4<sup>+</sup> T cells reactive to the immunizing antigen or peptide are essential for the disease to take its course.

## Concordance

Analysis of familial aggregation and disease concordance rates in monozygotic and dizygotic twins has also pointed to the genetic contribution to autoimmune disease. In MS, for example, the concordance rate for clinically expressed disease is 25–30 per cent for identical twins and less than 5 per cent for dizygotic twins (the same as in non-twin siblings)<sup>5</sup>. But because concordance in identical twins is far from 100 per cent, it is clear that other factors are involved.

What might those factors be? An infectious agent<sup>6</sup>? The nature of such a (possibly viral) 'trigger' is unknown, but several models have been proposed for how an infection could initiate disease in a susceptible person. Central to these models is the overwhelming evidence that auto-

immune and potentially disease-causing T cells are present in the normal peripheral T-cell repertoire.

These T cells are usually prevented from initiating disease by a variety of potential protective mechanisms which could be overcome by infection. For example, damage to the central nervous system could result in the release of sequestered antigens and cause T-cell sensitization. Alternatively, a peripheral infection and resultant immune response could trigger cross-reactive T cells with specificity for self antigens ('molecular mimicry'). Included in this latter mechanism, but different in concept, are powerful microbial antigens (superantigens) that target T cells expressing the same V $\beta$  and therefore large subsets of T cells that might include some with autoreactive specificities<sup>6</sup>.

So any type of activation of autoreactive T cells could circumvent protective mechanisms. Activated cells could acquire different trafficking and homing patterns, allowing them to enter target organs. Requirements for autoantigen presentation could also be affected, as could the pattern of cytokines produced. An EAE model, involving mice expressing transgenes for a TCR specific for MBP, bears on this point — the mice spontaneously developed neurological disease only if they were exposed to environmental pathogens<sup>7</sup>.

The work of Utz *et al.*<sup>1</sup> is the latest in a line of research into T cells in MS. This research has two main aims — characterization of the disease-causing T cells and their TCRs; and identification of the autoantigens being recognized.

The first task has proved to be especially formidable because of several difficulties. The T cells concerned may be affected by changes in the disease course and by medications used to treat the patient. Further, given the difficulty of obtaining and analysing samples of brain or cerebrospinal fluid, by necessity many investigators have chosen to work with blood T cells even though the relevance of these cells to what is going on in the brain is uncertain. Finally, even with the proper analysis of appropriate samples, it could be that non-specific T cells, recruited into the region of active inflammation<sup>8,9</sup>, would obscure autoreactive T cells in the brain and cerebrospinal fluid. So it is no surprise that the field is characterized by conflicting results over the TCR repertoire associated with MS<sup>10–16</sup>, results which at present defy rationalization.

As for identification of the autoantigens being recognized in MS, there are clear indications that MBP may be a

target<sup>4,15,17,18</sup>. Even if that is so, however, there is no evidence that it is the initiating autoantigen.

So much for the background to the new observations. In their studies on twins, Utz *et al.* find that the MBP-reactive TCR repertoire in blood has been altered in the twin with disease. It is, though, easier to explain how this shift in repertoire occurred as a result of disease rather than before disease.

## Evolution

For example, MBP-reactive T cells sequestered in the brain may not be detected when analysing blood. Perhaps more important, autoantigen responses may change during the course of disease. In EAE, an initial response to a dominant MBP peptide can lead to recognition and stimulation by new MBP determinants ('determinant spreading'), as well as responses to previously cryptic MBP determinants<sup>19</sup>. Such evolution could also influence the peripheral repertoire, as could the development of energy to some but not other determinants. Regulatory mechanisms that target TCR specific for MBP<sup>20</sup> might also be induced. If any of these processes do indeed occur, then longitudinal analysis of individual patients may show shifts in the MBP-specific repertoire over time. The new findings<sup>1</sup> might therefore provide a biological explanation for some of the variability observed in MBP-reactive repertoire in different studies.

A possible confounding factor, however, is that Utz *et al.* found that the twin with disease, compared to the one without, also expressed a different TCR repertoire in response to tetanus toxoid. This foreign antigen seems not to be related to the process of MS. Could it have been affected by chance? As brain damage occurs in MS, new self-antigens may be

1. Utz, U. *et al.* *Nature* **364**, 243–247 (1993).
2. Bernard, C. C. A. & de RosBo, N. K. *Curr. Opin. Immun.* **4**, 760–765 (1992).
3. Martin, R. *et al.* *Rev. Immun.* **10**, 153–187 (1992).
4. Zamvil, S. S. & Steinman, L. *Rev. Immun.* **8**, 579–621 (1990).
5. Ebers, G. *et al.* *New Engl. J. Med.* **315**, 1638–1642 (1986).
6. Kotzin, B. L., Leung, D. Y. M., Kappler, J. & Marrack, P. *Adv. Immun.* (in the press).
7. Goverman, J. *et al.* *Cell* **72**, 551–560 (1993).
8. Bell, R. B. *et al.* *J. Immun.* **150**, 4085–4092 (1993).
9. Karin, N., Szafer, F., Mitchell, D., Gold, D. P. & Steinman, L. *J. Immun.* **150**, 4116–4124 (1993).
10. Wucherpfennig, K. W. *et al.* *Science* **248**, 1016–1019 (1990).
11. Kotzin, B. L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 9161–9165 (1991).
12. Ben-Nun, A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 2466–2470 (1991).
13. Martin, R. *et al.* *J. exp. Med.* **173**, 19–24 (1991).
14. Oksenberg, J. R. *et al.* *Nature* **345**, 344–346 (1990).
15. Oksenberg, J. R. *et al.* *Nature* **362**, 68–70 (1993).
16. Wucherpfennig, K. W. *et al.* *J. exp. Med.* **175**, 993–1002 (1992).
17. Allegretta, M., Nicklas, J. A., Sriram, S. & Albertini, R. J. *Science* **247**, 718–721 (1990).
18. Chou, Y. K. *et al.* *J. Neuroimmun.* **38**, 105–114 (1992).
19. Lehmann, P. V., Sercarz, E. E., Forsthuber, T., Dayan, C. M. & Gammon, G. *Immun. Today* **14**, 203–208 (1993).
20. Offner, H., Hashim, G. A. & Vandenbark, A. A. *Science* **251**, 430–432 (1991).

recognized and targeted<sup>19</sup>, and one of them might just cross-react with tetanus toxoid. Or is it possible that a major shift in repertoire has occurred in patients independent of the target autoantigens and before onset of disease? This shift could represent the non-genetic element that explains discordance. Such a shift, if induced by an environmental agent, would require some type of truly super antigen, or several different antigens. Utz *et al.* report only  $V\alpha$  usage, so it is difficult to tell whether a  $V\beta$ -specific superantigen, as currently known<sup>6</sup>, might be involved.

In all, we now have one more complexity to take into account in studies

of autoimmune disease: clearly, the responding TCR repertoire in patients with chronic disease is affected by more than just the genetic background. The next steps will be to examine more sets of twins and more antigens, and to include analysis of  $V\beta$  expression. It will also be interesting to see if longitudinal studies in individual patients show TCR repertoire shifts, and if similar phenomena crop up in other chronic autoimmune diseases. □

*Brian L. Kotzin is in the Division of Basic Sciences, National Jewish Center for Immunology and Respiratory Medicine, Denver, Colorado 80206, USA.*

a salt dome in West Kazakhstan. Data on this experiment (which was not clandestine) have recently been made available; it was observed teleseismically, both inside and outside the USSR, and had a magnitude of about 4.1 (Sykes). The signal amplitudes were only about a factor of 10 lower than for a tamped shot of similar yield.

But even if decoupling is difficult (because of the challenges of building large, stable underground cavities, and secretly carrying out a sophisticated nuclear test that does not leak radioactive evidence), it still has some influence on assessments of monitoring capability.

With new seismic instrumentation, we can probably be confident that global monitoring would detect any explosion down to about magnitude 4. Without evasion this magnitude corresponds to about 1 kilotonne. With evasion (albeit by methods that some consider impractical, and that would be risky in the light of all the monitoring techniques that could be brought to bear), magnitude 4 might correspond to about 10 kilotonnes.

Below magnitude 4, seismic monitoring becomes difficult. Signal-to-noise ratios are lower; there are typically fewer seismic stations recording each small event; and there are large numbers of natural small earthquakes and chemical explosions, which must be identified as non-nuclear. New ways of discriminating between explosions and earthquakes are being developed (H. Patton, Lawrence Livermore National Laboratory), and spectral techniques can often now distinguish between single shots and the typical ripple-fired chemical blasts used in quarries and mines (B. Stump, Southern Methodist Univ.). But the main improvements in monitoring capability that will be needed to support a new test-ban treaty

## SEISMOLOGY

# Testing the test-ban treaty

*Paul G. Richards*

As serious efforts to negotiate multilateral limitations on nuclear testing may resume this year after a 12-year hiatus, it is now necessary to look at the progress made on verification, the problems that remain and the plans for explosion monitoring needed to support a new treaty. The first message emerging from two recent meetings\* was that seismic methods can be used to monitor nuclear explosions down to yields far smaller than the threshold of 150 kilotonne high explosive equivalent set by the current bilateral treaty between the United States and Russia on nuclear testing.

There remains a grey area between about 1 and 10 kilotonnes where experts disagree on verification capability. And at 1 kilotonne and below (a level at which some weapons designers claim there is a limited capability for testing and development of nuclear weapons or their components), verification cannot be assured with high confidence, even if plans for new seismometers are developed on a global basis. A clash is therefore brewing on a major question of arms control policy between those who advocate a multilateral, comprehensive test-ban treaty both as a restraint on the nuclear states and in furtherance of the goals of the Nonproliferation Treaty, and those who oppose such a treaty on grounds either of a perceived need to continue testing, or as a matter of principle that an unverifiable (comprehensive) treaty is bad policy.

Since nuclear testing went underground in the 1960s, seismology has provided the most important methods for detecting and locating explosions, and distinguishing them from earthquakes. A problem has always been the possibility that a nuclear

explosion might be executed in an underground cavity, thus decoupling or muffling the seismic signals to make them look like a much smaller shot carried out under typical (tamped) conditions. In the early 1960s, it was proposed that an explosion as large as 300 kilotonnes could be decoupled to look like a 1-kilotonne tamped explosion. But in fact, decoupled explosions are less of a problem for two reasons (L. R. Sykes, Columbia Univ.): cavities for effective decoupling are too large for practical construction for more than about 10 kilotonnes; and the decoupling is far less efficient than was originally thought. Evidence for this comes from a decoupled explosion carried out by the then Soviet Union on 29 March 1976, when about 10 kilotonne high explosive equivalent was exploded in the cavity of a previous 64-kilotonne shot in



Surface preparations for an underground test of a nuclear bomb at the Nevada test site.

\*All-day session at American Geophysical Union meeting, Baltimore, 27 May 1993; and the Fifth Annual Workshop of the IRIS Consortium, Waikoloa, Hawaii, 10–14 June 1993.



## Call for extended moratorium

**PRESIDENT Clinton on 3 July called on the other nuclear powers to join the United States in a moratorium on nuclear testing that he has ordered to be extended for at least 15 months. Saying that "If these nations will now join us in observing this moratorium, we will be in the strongest possible position to negotiate a comprehensive test ban", he also warned that the United States would move quickly to conduct nuclear tests if there was any resumption of testing by another country before October 1994. British testing in recent years has been conducted only at the main United States test site in Nevada, and President Yeltsin has expressed Russian support for a comprehensive ban, so the main question now is what France and China will do. France has been under a recent moratorium, but has a new government. China carried out two nuclear tests in 1992 (one of them, at about 600 kilotonnes, the largest underground nuclear explosion by far since the mid-1970s). A comprehensive test ban may be closer than at any time since the efforts of Kennedy, Khrushchev and Macmillan in the early 1960s — but this Pandora's box will not close easily.**

P.G.R.

will stem directly from deployment of new seismic instrumentation.

In recent years, seismic stations of a new type have been installed around the world for purposes of geophysical research (G. van der Vink, IRIS consortium). Their distinguishing characteristics include broadband digital recording, and the ability to send data from remote sites, by modem and/or Internet connection, on request from any interested user. The global network has been a major effort of the international seismological research community, and its performance will, it is hoped, continue to improve. By combining the data from current and planned stations, it is estimated that all seismic events in Eurasia above about magnitude 3.8 could be detected with high confidence (J. Claassen, Sandia National Laboratory). The network performs even better than this in Central Asia, in large part because IRIS (a group of about 80 research institutions) has recently installed state-of-the-art seismometers and recording systems on territory of the former Soviet Union.

But the research community's network alone should not bear the burden of monitoring compliance with a weighty new treaty. If an event such as a nuclear test by a new state is very unlikely — but would lead to sweeping policy changes if it occurred — then the United States government, for one, would need a very reliable monitoring system (L. Turnbull, Central Intelligence Agency). The Advanced Research Projects Agency (ARPA) of the United States Defense Department in 1992 proposed the need to monitor all countries of high interest for signals of magnitude 2.5 and above (representing 1 kilotonne fully decoupled); to monitor globally to about magnitude 3–3.5 (about 10 kilotonnes fully decoupled); and to be able to locate seismic sources to within 10 kilometres "in order to bring other assets to bear" (Turnbull). These are demanding requirements, and some might say unattainable, as more than 50,000 earthquakes and several thousand

chemical explosions each year give signals above magnitude 3. In practice, monitoring below magnitude 3.5 is needed only in regions within which decoupling is plausible. Which are these regions? This question will influence the new effort needed in building a monitoring network. ARPA has been a strong proponent of building arrays of seismometers (typically 10 or more instruments deployed in an area a few kilometres across) to achieve better signal-to-noise ratios than single stations. A subcommittee of the United Nations in Geneva, known as the Group of Scientific Experts (GSE), has conducted technical tests funded extensively by ARPA to exchange seismic data between arrays and single stations in different countries. The GSE has plans (R. Alewine, ARPA) for a new international seismic centre that would continuously receive data from a global network of about 50 stations and arrays, backed up by a secondary network of stations from which data could be requested as needed, for detection and location of earthquakes and explosions.

The Clinton administration has yet to decide whether the GSE approach should receive support. It would be rational first for a decision to be made on the monitoring capability required for various countries, as a basis for negotiating new limitations on nuclear testing. If an effective end to nuclear testing is desired, but a comprehensive ban is deemed unachievable because small explosions might go undetected, then one approach would be a treaty banning all explosions at and above a certain size, together with an indefinite moratorium on all nuclear explosions below that size. Such an approach was developed in 1960–62 by the United States and the Soviet Union; with a treaty threshold set initially at, say, magnitude 4, perhaps it can succeed in the 1990s. □

Paul G. Richards is at the Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA.

## RÉSUMÉ

### Chiral practice

CARBON chemistry developed a new twist when the fullerene  $C_{76}$  turned out to exist in two enantiomeric forms (R. Ettl *et al.* *Nature* **353**, 149–153; 1991). But how might one separate the two? Lacking a nanoscale equivalent of Pasteur's tweezers, J. M. Hawkins and A. Meyer (*Science* **260**, 1918–1920; 1993) have taken the chemical approach of tagging the two twisted carbon cages with a chiral ligand. Each enantiomer of the ligand has a preference for just one of the two  $C_{76}$  stereoisomers. So separating the tagged and untagged products, and then removing the ligand from the tagged variant, yields 97 per cent pure, optically active  $C_{76}$  fullerenes.

### Blood and guts

To reproduce its kind the female mosquito needs to drink blood: two blood meals suffice to fuel oogenesis. But the blood proteins must first be broken down by gut enzymes, among which, as H.-M. Müller *et al.* (*EMBO J.* **12**, 2891–2900; 1993) have now shown, a pair of serine proteases, with the sequence attributes of trypsins, are prominent. Other trypsin-like sequences could be identified in the genome, but the two that Müller *et al.* have expressed in *Escherichia coli* do indeed digest blood proteins *in vitro*. Moreover their messenger RNAs appear in the mosquito gut only after the second blood meal. The interesting questions now are how feeding provokes transcription of the trypsin genes, and whether an antitrypsin for example, administered to the host, could spell death to the mosquito with more certitude than a folded newspaper.

### Looking back

SOME people just have to be different. Flouting the custom of using a microscope to investigate a sample, L. Montelius and J. O. Tegenfeldt have employed a cleverly designed sample surface to gaze at the tip of their atomic force microscope (*Appl. Phys. Lett.* **62**, 2628–2630; 1993). In scanning force microscopy, the shape of the tip is convoluted with that of the sample surface in the image, and the effects can be hard to remove. Montelius and Tegenfeldt get around this by etching an indium phosphide surface into tiny columns which look, in the scanning electron microscope, like an army of miniature villi. Being so much thinner than the tip, the columns act in effect as an array of delta functions. The image produced is a corresponding array of pyramidal tip shapes, one for each column, and broadened only slightly by their 40–50 nm diameter. Molecules attached to the microscope tip itself could, say the authors, be imaged simultaneously by dozens of the columns, giving a painless increase in the statistical significance of each scan.

# The road not taken

Marc Montminy

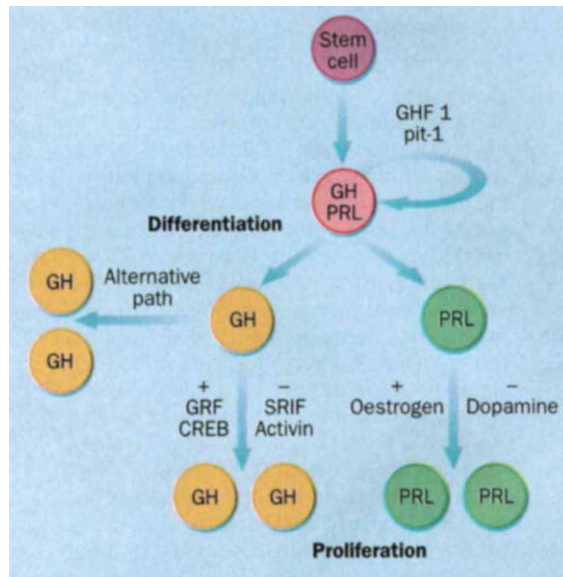
SOMEONE once pointed out that a dwarf standing on the shoulders of a giant can see further than the giant himself. Well, dwarfs alone — strains of dwarf mice, as it happens — now give us an unexpectedly clear view of the development of the pituitary gland. The work concerned centres on growth-hormone-producing cells called somatotrophs, found in the pituitary, which have become a focal point for researchers interested in how extracellular signals influence cell ontogeny. Two reports — one in *Nature Genetics*<sup>1</sup> and the other on page 208 of this issue<sup>2</sup> — show that an inherited mutation in the growth-hormone-releasing factor (GRF) receptor gene leads to dwarfism and somatotroph hypoplasia (decreased numbers of growth-hormone-producing cells) in the *little* mouse.

The idea that GRF might function not only as a hypothalamic-releasing hormone, but also as a trophic factor, began with clinical reports of patients with GRF-secreting tumours who showed elevated levels of circulating growth hormone and markedly enlarged anterior pituitaries replete with hyperplastic (abnormally proliferating) somatotrophs<sup>3,4</sup>. Subsequent studies on primary cultures of rat anterior pituitary cells demonstrated that GRF could induce both growth-hormone transcription<sup>5</sup> and somatotroph proliferation<sup>6</sup> by means of a cyclic AMP-dependent mechanism.

An unexpected twist in this story came when researchers found that almost half of pituitary tumours from acromegalic patients had markedly increased levels of cAMP, owing to point mutations in the  $\alpha_s$  subunit of the heterotrimeric  $G_s$  protein which led to constitutive activation of adenylyl cyclase<sup>7</sup>. Indeed, transgenic mice expressing a cholera toxin gene under control of the growth-hormone promoter also developed somatotroph hyperplasia<sup>8</sup> suggesting that chronic activation of the cAMP pathway was sufficient to promote hyperplasia and oncogenic transformation there. Moreover, transgenic mice overexpressing an inactive mutant of the cAMP-responsive transcription factor CREB, under control of the growth-hormone promoter, showed selective depletion of somatotroph cells<sup>9</sup>.

Where, in the ontogeny of the anterior pituitary, might GRF act through cAMP to amplify or maintain the somatotroph lineage? Cellular lineage studies (see

figure) in transgenic mice expressing thymidine kinase<sup>10</sup> or diphtheria toxin<sup>11</sup> genes under control of the growth-hormone promoter pointed to the existence of a common 'somatomammotroph' stem cell precursor. And studies in Snell and Jackson<sup>12</sup> dwarf strains showed that these mice have hypoplastic pituitaries deficient in somatotroph, lactotroph and thyrotroph cell lineages due to missense



Population of the pituitary by somatotrophs and lactotrophs. Positive inducers for each path are shown under (+) sign and inhibitors indicated below (–) sign. The looped arrow next to the pluripotent somatomammotroph (GH/PRL) indicates a capacity for self-regeneration. GH, growth hormone; PRL, prolactin; SRIF, somatostatin.

mutations in the POU-domain transcription factor pit1/GHF. As antisense oligonucleotides to pit1/GHF directly inhibit proliferation of a somatotroph cell line GC, these results would suggest that pit1/GHF also plays an important mitogenic role<sup>13</sup>.

The *little* (*lit*) mice manifest selective somatotroph hypoplasia because of a missense mutation in the ligand-binding domain of the GRF receptor<sup>1,2</sup>. Somatotroph numbers in fetal *lit/lit* pituitaries are almost indistinguishable from their wild-type counterparts, with hypoplasia of growth-hormone producing cells becoming evident only postnatally in the caudo-medial extent of the gland<sup>2</sup>. Remarkably, the anterolateral segment of anterior *lit/lit* pituitary appears to maintain normal numbers of somatotrophs which, as the authors suggest<sup>2</sup>, may signify a zone of self-renewing somatotroph stem cells. What remains to be shown is whether these cells are truly proliferating or whether they might be post-mitotic, like

most lactotrophs. In either case, the sparing of somatotrophs in *lit/lit* mice shows that GRF is not absolutely essential for generation of all pituitary somatotrophs. And the presence of a residual GRF-independent somatotroph population argues for the presence of some growth factor or, conversely, for the absence of an inhibitor within the anterolateral segment of the pituitary.

Work with the synthetic growth-hormone-releasing peptide-6 (GHRP-6) provides strong evidence for the presence of a second receptor on somatotrophs which is distinct from GRFR<sup>14</sup>. GHRP-6 apparently signals through the phosphoinositol pathway and can synergize with GRF to promote growth hormone release. In the absence of GRF (as in *lit/lit* mice) anterolateral pituitary cells may respond to this alternative GHRP-6 sensitive pathway and thus proliferate normally.

Sparing of anterolateral somatotrophs may also reflect loss of negative regulation by hypothalamic or pituitary factors. The inhibitory effects of somatostatin on somatotroph proliferation, for example, are well characterized and may explain, in part, some of the phenotypic changes observed in *lit/lit* pituitaries. Anterolateral stem cells may escape negative regulation either because they do not express somatostatin receptors or because they are not exposed to inhibitory concentrations of this peptide.

Somatotroph-inhibitory factors may also arise from within the pituitary itself. The activins, for example, function as potent inhibitors of somatotroph proliferation and growth-hormone transcription<sup>15</sup>. And the demonstrated expression of activin polypeptides within pituitary gonadotrophs implies that this paracrine mechanism of inhibition is at least plausible.

Several questions remain. What, for example, are the inductive signals that

1. Mayo, K. E. *et al.* *Nature Genet.* **4**, 227–232 (1993).
2. Lin, S.-C. *et al.* *Nature* **364**, 208–213 (1993).
3. Rivier, J., Spiess, J., Thorner, M. & Vale, W. *Nature* **300**, 276–279 (1982).
4. Guillemin, R. *et al.* *Science* **218**, 585–587 (1982).
5. Barinaga, M. *et al.* *Nature* **306**, 84–85 (1983).
6. Billestrup, N., Swanson, L. W. & Vale, W. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6854–6857 (1986).
7. Landis, C. A. *et al.* *Nature* **340**, 692–696 (1989).
8. Burton, F. H., Hasel, K. W., Bloom, F. E. & Sutcliffe, J. G. *Nature* **350**, 74–77 (1991).
9. Struthers, R. S., Vale, W. W., Arias, C., Sawchenko, P. E. & Montminy, M. R. *Nature* **350**, 622–624 (1991).
10. Borrelli, E., Heyman, R. A., Arias, C., Sawchenko, P. E. & Evans, R. M. *Nature* **339**, 538–541 (1989).
11. Behringer, R. R., Mathews, L. S., Palmiter, R. D. & Brinster, R. L. *Genes Dev.* **2**, 435–461 (1988).
12. Li, S. *et al.* *Nature* **347**, 528–533 (1990).
13. Castrillo, L., Theill, L. & Karin, M. *Science* **253**, 197–199 (1991).
14. Smith, R. G. *et al.* *Science* **260**, 1640–1643 (1993).
15. Struthers, R. S., Gaddy-Kurten, D. & Vale, W. W. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11451–11455 (1992).



allow GRF receptors to be expressed on somatotrophs but not on lactotrophs? Clearly both lactotrophs and somatotrophs express pit1/GHF, which is shown here to regulate GRF-receptor promoter activity. But lactotrophs do not express the GRF receptor, suggesting that other regulatory factors must be produced after the bifurcation of lactotrophs and somatotrophs. A related question concerns the acquisition of cAMP-responsive proliferation by somatotrophs but not lactotrophs postnatally. The latter problem is particularly intriguing in view of the fact that although cAMP is expressed ubiquitously, it serves as a mitogen only in specific

endocrine cell types such as the somatotroph.

Whatever the answers, the road to understanding pituitary development promises to be an exciting one. In this regard, it is noteworthy that the two groups reporting on the *little mouse*<sup>1,2</sup> have longstanding interests in GRF and share a common origin scientifically — thus providing yet another example of science imitating life. □

Marc Montminy is in The Clayton Foundation Laboratories for Peptide Biology, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.

## SEISMIC TOMOGRAPHY

# Deep thoughts on the mantle

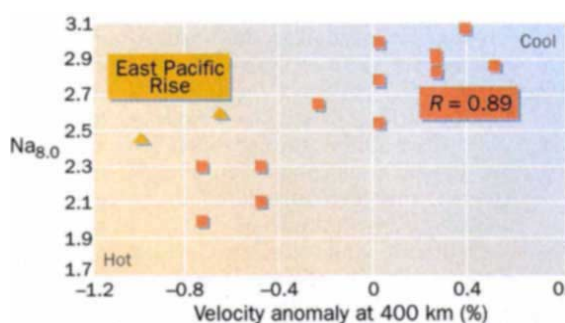
Charles H. Langmuir

THE distributions of temperature and composition in the Earth's mantle are questions being tackled from many directions, and lines of attack from separate disciplines often proceed in isolation. When evidence from two disciplines converges, the results from each become strengthened by the linkage with independent data sets and reasoning. Humler, Thiriot and Montagnier<sup>1</sup>, on page 225 of this issue, present such a linkage between ocean ridge basalt compositions and the velocity anomalies deduced from seismic tomography.

Petrologists have made use of basalt compositions from rocks dredged from the sea floor to evaluate temperature variations in the mantle beneath ocean ridges<sup>2-6</sup> and have concluded that variations of more than 200 °C are required to explain the observed fluctuations in composition. Such temperature differences would also explain the variations in crustal thickness and depth of the ocean ridge<sup>2</sup>. Seismic tomography can also be used, if indirectly, to derive the temperature distribution. Lateral variations in the shear wave velocity can be interpreted as resulting from regional differences in mantle temperature, with slower velocities arising from hotter mantle. Humler and co-workers suggest that these very different methods yield similar results for the temperature distribution in the mantle beneath ocean ridges. Using Zhang and Tanimoto's<sup>7</sup> inversion of the seismic data, they show that if the velocity variations at a depth of 110 kilometres are caused purely by temperature anomalies, then the range of temperature estimates from the petrological results<sup>2,4</sup> and tomographic results<sup>7</sup> agree.

At shallower depths, seismic velocities would also be strongly influenced by the presence of small amounts of melt in the mantle beneath the ridge.

There are, however, several perplexing



Comparison of  $\text{Na}_2\text{O}$  contents of basalts normalized to 8 per cent  $\text{MgO}$  (this is written  $\text{Na}_{8.0}$ ), with the velocity anomalies at a depth of 400 kilometres obtained from the model of Su *et al.*<sup>10</sup>. The  $\text{Na}_{8.0}$  values are averages over 15-degree windows of latitude, using regionally averaged data from refs 2 and 4. The standard deviations in  $\text{Na}_{8.0}$  are less than 0.2 in each case. The velocity anomalies are read directly from Fig. 3e of Su *et al.*<sup>10</sup> and their uncertainties are about  $\pm 0.2$ . The correlation coefficient for the data, excluding the two points for the East Pacific Rise (orange triangles), is  $R = 0.89$ . If those two points are included,  $R = 0.81$ .

aspects to the relationships between the petrological and seismic data. One is the anomalous data points: some geographical regions from other petrological compilations<sup>2</sup> are missing from the data set<sup>1</sup>, and others do not fit. For example, the shallow (and petrologically very hot) Kolbeinsey ridge north of Iceland is seismically fast instead of slow. Indeed, the high velocities beneath Iceland and surrounding ridges have been the impediment to previous attempts to find correlations between tomographic and petrological results. As this problem diminishes with more recent tomographic results, it seems that the apparent lack of agreement

may be one of resolution and spreading rate rather than substance. The old, cold lithospheres beneath Scandinavia and Greenland, which border the narrow ocean basin around Iceland, may strongly influence tomographic results based on low-order spherical harmonics, which have little resolving power at scales less than thousands of kilometres. And the diminishing spreading rate north of Iceland leads to colder mantle close to the ridge axis.

A more serious dilemma is that the correlation peters out at depths greater than 110 kilometres. As hot rock upwells and the pressure on it decreases, it will expand adiabatically and melt when it reaches the solidus appropriate for its composition and temperature. The depth of 110 kilometres, where the best correlations are found, is beneath the dry mantle solidus for most ridges, so it could reflect variations in mantle temperature rather than melt content. But the physical and petrological models would require that these temperature variations persist with increasing depth. If they do, the correlations should be seen at all depths greater than 110 kilometres. Yet the seismic data show that anomalies decrease sharply with depth, and the correlations seen by Humler and co-workers disappear rather than persist. This is a difficult result to explain.

In fact, one could argue that it means that temperature anomalies are not the explanation. An alternative solution would be the presence of different amounts of melt, caused by variable volatile content in the mantle. Volatile-rich mantle would begin to melt while still at greater depths, 150 kilometres or more, and would then have more melt at 110 kilometres than drier regions that might not even have begun to melt significantly at that depth. In this case, the seismic velocity differences would be caused by the presence of different amounts of melt rather than temperature, and the basalt compositional variations might result from differences in mantle composition.

These results will fuel the controversies that have arisen for both petrological and seismic data sets. Albaredo<sup>8</sup> and Shen and Forsyth<sup>9</sup> have argued that the petrological data can be explained by heterogeneity in the mantle's composition rather than by temperature variations. And although Zhang and Tanimoto<sup>7</sup> have argued that the velocity anomalies beneath ocean ridges are shallow (generally less than 150 kilometres below the sea floor), Su *et al.*<sup>10</sup> assert that there are serious problems with the Zhang and Tanimoto inversion, and that velocity variations associated with

ridges run much deeper, to as much as 600 kilometres.

We can investigate further by comparing the petrological data with the deeper seismic velocity anomalies reported by Su *et al.*<sup>10</sup>. Because their seismic model has a resolution of about 2,000 kilometres (harmonic degree 13), I have averaged the basalt data over roughly 15-degree intervals of latitude, and compared the data to the velocity anomalies at a depth of 400 kilometres.

The critical parameter for the basalt data is  $\text{Na}_{8,0}$  (see figure), which is inversely proportional to the extent of melting, and hence inversely proportional to mantle temperature. At this scale, there is a good correlation between the petrological and seismic results even at a depth of 400 kilometres although the two data points from the fast-spreading East Pacific Rise are still slightly slow, as they are at all shallower depths in the model of Su and collaborators<sup>10</sup>.

The persistence of a correlation down to 400 kilometres contrasts with the results found by Humler *et al.*, yet resolves the difficulty that their results imply. The correlation seen in the figure is strongly suggestive of deep-seated variations in mantle temperature and would be fully consistent with the petrological results. Variations in the seismic velocity at 400 kilometres are similar to those that Humler *et al.* report at shallower levels, so they give the same range of seismically inferred temperature variations as those inferred from the petrological results.

If the Zhang and Tanimoto<sup>7</sup> inversion gives better resolution at shallow levels, whereas the inversion of Su *et al.*<sup>10</sup> gives more reliable results for regional differences at deeper levels, then the agreement between the two petrological and tomographic data sets would indeed be consistent with large-scale lateral variations in the temperature of the upper mantle. As the resolution of both petrological and seismic data improves, so should our understanding. Further developments will be warmly welcomed. □

Charles H. Langmuir is at the Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA.

# The oceans and global warming

Andrew J. Weaver

In the debate on global warming, the role of the oceans is one of the key questions. Forecasts for the next century have suggested that the oceans will help to curb global warming by absorbing vast quantities of heat. Manabe and Stouffer<sup>1</sup>, on page 215 of this issue, take a longer view. If carbon dioxide emissions were to continue until the atmospheric concentration doubled, they say, the North Atlantic's 'conveyor belt' for heat transport would weaken but would recover, given time; if concentrations quadrupled, this recovery might never occur.

Their conclusions should not be lightly dismissed, for the long-term warming of surface air temperature predicted with  $\text{CO}_2$  doubled is an average of 3.5 °C worldwide, accompanied by at least a 1-m rise in sea level. If we fail to curb greenhouse gas emissions before concentrations soar to four times present values, our descendants several centuries from now could face an average temperature rise of 7 °C and sea levels at least 1.8 m higher than today.

In the present climate, the North Atlantic and Southern oceans are the two regions of deep water formation where warmer surface waters are converted to colder deep waters through intense heat loss to the atmosphere. In the North Atlantic, this high-latitude cooling combined with low-latitude heating accelerates a thermohaline circulation with poleward flow at the surface. On the other hand, net high-latitude precipitation, runoff and ice melt (feeding in fresh and thus less dense water) together with low-latitude evaporation tend to drive the circulation in the opposite direction. In today's North Atlantic the first effect — the thermal forcing — dominates over the second (haline) forcing. The importance of the thermohaline circulation is that it transports heat from equatorial regions to the poles, acting as a large-scale conveyor system.

There has already been work on the initial response of the North Atlantic conveyor to mounting concentrations of atmospheric  $\text{CO}_2$ , using climate models that included the interactions between oceans and atmosphere. The results indicated that it would weaken initially in response to an increase corresponding to roughly what will be seen if emissions continue at the present rate (this is known as the 'Business As Usual' scenario A of the Intergovernmental Panel on Climate Change<sup>4</sup>). With the weakened conveyor comes a reduction in northward oceanic heat transport and the commencement of slow oceanic heat absorption.

Manabe and Stouffer now look further

ahead. In the first of their model experiments,  $\text{CO}_2$  was increased at a rate of 1 per cent per year for 70 years (doubling the present concentration) and then held constant for the rest of a 500-year integration. In the second, the  $\text{CO}_2$  increase continued for a further 70 years until the concentration had quadrupled. Initial predictions were similar to those seen earlier — a weakened North Atlantic conveyor, and a rise in average global surface air temperature at a rate of 3.5 °C per century. After the first 70 years, once the two experiments had diverged, their responses became very different. In the first case, the thermohaline circulation began to reintensify around year 150 until it was completely re-established. In the second case it remained weak.

These two experiments emphasize the delicate balance between thermal and freshwater flux forcing in driving the present-day North Atlantic conveyor. In the  $\text{CO}_2$ -doubling case, subsurface warming of the ocean (stronger when the circulation is weak) acts to increase the Equator-to-pole density contrast and hence the circulation, because warm water has a larger coefficient of expansion than cold water. In the  $\text{CO}_2$ -quadrupling case, greater high-latitude precipitation resulting from increased water-vapour transport, together with increased melting of sea ice, weakens the density contrast and the circulation. Although the deep ocean may not have equilibrated fully in the second case, these results suggest that different equilibrium climate states<sup>5</sup> can be realized depending on how long the atmospheric  $\text{CO}_2$  continues to increase.

Manabe and Bryan's earlier work with an idealized single-hemisphere model of the ocean and atmosphere<sup>6</sup> (containing three continents and three oceans each of 60° width) suggests that the total poleward heat transport (in the ocean and atmosphere) may remain fairly constant for climatic equilibria reached when  $\text{CO}_2$  is multiplied two-, four- or even eightfold. The reason is that the radiation balance at the top of the atmosphere remains largely unchanged.

The enhanced high-latitude warming predicted by Manabe and Stouffer would tend to reduce the poleward atmospheric transport of sensible heat (the heat advected by the winds). If the weakened thermohaline circulation is re-established and carries its usual quantities of heat, the reduced atmospheric transport of sensible heat is compensated for by a slight increase in atmospheric transport of latent heat (associated with the meridional transport of water vapour). But in the  $\text{CO}_2$ -quadrupling case, where the circula-

1. Humler, E., Thiriot, J. L. & Montagnier, J. P., *Nature* **364**, 225–228 (1993).
2. Klein, E. & Langmuir, C. J. *geophys. Res.* **92**, 8089–8115 (1987).
3. McKenzie, D. & Bickle, M. J. *J. Petrol.* **29**, 625–679 (1988).
4. Langmuir, C., Klein, E. & Plank, T. *AGU Monogr.* **71**, 183–280 (1992).
5. Kinzler, R. J. & Grove, T. L. *J. geophys. Res.* **97**, 6907–6926 (1992).
6. Dick, H. J., Fisher, R. L. & Bryan, W. B. *Earth planet. Sci. Lett.* **69**, 88–106 (1984).
7. Zhang, Y. S. & Tanimoto, T. *J. geophys. Res.* **98**, 9793–9823 (1993).
8. Albareda, F. J. *geophys. Res.* **97**, 10997–11008 (1992).
9. Shen, Y. & Forsyth, D. W. *Eos* **74**, 283 (1993).
10. Su, W.-J., Woodward, R. L. & Dziewonski, A. M. *Nature* **360**, 149–152 (1992).



tion remains weak, this latent-heat transport in the atmosphere has to be much greater. This is internally self-consistent with the collapsed thermohaline circulation, as the warmer atmosphere holds more moisture and there is more high-latitude precipitation which, together with enhanced melting of sea ice, creates a freshwater cap over the regions of deep water formation.

By systematically comparing the results from fully coupled atmosphere–ocean models, such as that of Manabe and Stouffer, with those of atmospheric models in which only an ocean surface layer is considered, we can begin to understand the importance of the ocean in climate change. In the short term, the ocean acts to reduce the effects of global warming. This is particularly noticeable in the high-latitude regions of the North Atlantic, where the thermohaline circulation is weakened, and in the Southern Hemisphere, where effective deep mixing of heat anomalies markedly reduces the warming. Both fully coupled atmosphere–ocean models<sup>2,3</sup> and atmospheric models with only an ocean surface layer<sup>2,7</sup> give similar responses in regions where the thermohaline circulation is sluggish (such as the Pacific and Indian oceans).

Both classes of model further suggest that greater high-latitude warming is due to the reduction of sea-ice cover and the accompanying decrease in surface albedo. Regions of more active thermohaline circulation (Southern and Atlantic oceans) are treated better in the fully coupled models. Manabe and Stouffer have shown that it is these regions where the oceanic response can be dramatically different depending on how long CO<sub>2</sub> continues to increase.

We can now extend our understanding to regional climate changes associated with a weakened or active circulation. With the circulation collapsed, as it seems it would with CO<sub>2</sub> quadrupled, the warming over the North Atlantic and hence Europe is smaller than if it had been active, although the enhanced meridional water-vapour transport means more rain. In the CO<sub>2</sub>-doubling case, the established thermohaline circulation acts to amplify the warming of the North Atlantic and hence Europe.

Perhaps the ocean's role in global

warming is becoming clearer, but there are still large uncertainties on the atmospheric side. Water vapour and cloud feedbacks<sup>8</sup> are often poorly resolved in coarse-resolution climate models, and the basic physics is not yet clear. Furthermore, these models do not, as yet, explicitly incorporate oceanic and terrestrial uptake or release of anthropogenic and natural greenhouse gases; and artificial flux corrections must still be used<sup>5</sup>. Even though the most sophisticated cli-

mate models predict global warming in response to increasing greenhouse gases, we are still searching for the signal in our observational network amid a background of natural climatic variability. Climate scientists look likely to have their hands full for the next few years. □

Andrew J. Weaver is at the School of Earth and Ocean Sciences, University of Victoria, PO Box 1700, Victoria, British Columbia V8W 2Y2, Canada.

## NEUROPSYCHOLOGY

# Imagine only the half of it

John C. Marshall and Peter W. Halligan

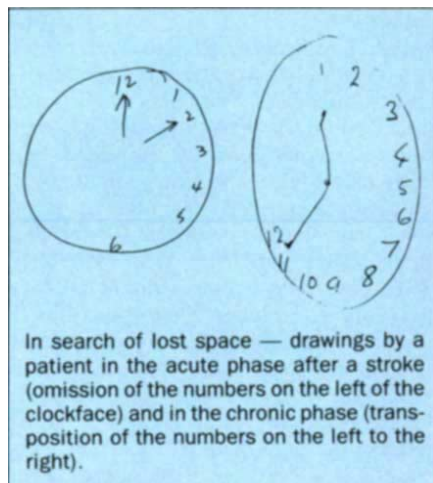
THE idea that vision and visual imagery draw upon much the same central mechanism takes a knock on page 235 of this issue<sup>1</sup>. There, Guariglia and colleagues report studies on a patient who suffered a stroke, and (it turned out) injury to the right frontal lobe of the brain. Against expectation, the patient is

quired. More convincing evidence that visual neglect is not, strictly speaking, visual was provided some years ago in a seminal paper by Bisiach and Luzzatti<sup>3</sup>, and the idea is now given an unexpected twist by Guariglia *et al.*

The patients studied by Bisiach and Luzzatti were asked to describe, from memory, the Piazza del Duomo in Milan, where they had lived for many years. From an (imaginary) vantage point with their back to the front door of the Duomo, the patients reported far more details from the right side of the piazza than from the left. But when requested to imagine themselves on the other side of the Piazza facing the Duomo, they again described more features on the (new) right (previously left) side of the piazza. This pattern of 'representational' neglect has subsequently been found in many patients who also show left neglect on tasks where sensory information is presented across the visual field.

By contrast, the patient of Guariglia *et al.* showed no neglect in personal space (for the left side of his body), in peripersonal space (for objects within reaching distance) or extrapersonal space (for objects beyond arm's length). Yet, on imagery tasks comparable to those used by Bisiach and Luzzatti, his verbal descriptions showed severe neglect of geographical features on the left of Roman piazzas (where 'left' again means left from two opposite vantage points on separate trials). There were both omissions of features on the left and transpositions of features from the left to the right side of the piazzas. No previously reported patient with 'imaginal' neglect has failed to show visual neglect.

A single dissociation between imagery and sensation nonetheless raises the issue of whether a full double dissociation can occur. That is, are there patients who show left neglect in situations where a visual stimulus is presented but have no neglect on imaginal tasks? We know of no study strictly comparable to that of



deficient in carrying out tasks involving imagery in the left visual field but not in other aspects of visuo-spatial cognition involving that space. This is a phenomenon that will give psychologists and neurologists pause for thought.

In the so-called hemispatial neglect syndromes, patients who (typically) have suffered severe damage to the right temporo-parietal part of the brain seem to live in "only half the world"<sup>2</sup>. Visual objects in contralesional (left) space fail to be detected and responded to even in free vision with a full range of head and eye movements. Although many (but not all) of these patients are also blind in the left field, this condition, known as hemianopia, is not the primary determinant of their problem; patients who only have a hemianopia quickly learn to compensate by making more extensive eye and head movements than would normally be re-

1. Manabe, S. & Stouffer, R. J. *Nature* **364**, 215–218 (1993).
2. Manabe, S., Stouffer, R. J., Spelman, M. J. & Bryan, K. *J. Clim.* **4**, 785–818 (1991).
3. Cubasch, U. *et al.* *Clim. Dyn.* **8**, 55–69 (1992).
4. Intergovernmental Panel on Climate Change *The IPCC Scientific Assessment* 366 (WMO-UNEP, Cambridge Univ. Press, 1990).
5. Manabe, S. & Stouffer, R. J. *J. Clim.* **1**, 841–866 (1988).
6. Manabe, S. & Bryan, K. *J. geophys. Res.* **90**, 11689–11707 (1985).
7. Boer, G. J., McFarlane, N. A. & Lazare, M. *J. Clim.* **5**, 1045–1077 (1992).
8. Intergovernmental Panel on Climate Change *The Supplementary Report to the IPCC Scientific Assessment* 200 (WMO-UNEP, Cambridge Univ. Press, 1992).

## Chaos for oscillating voles

It's a tough life, being a vole in northern Fennoscandia. If the harsh climate and the ever-present peril of being eaten by a weasel *Mustela nivalis* (pictured here) weren't enough, there is now the Lyapunov exponent to cope with. For as Ilkka Hanski and colleagues show on page 232 of this issue, the life of the vole is more than



Jussi Murtosaari

just difficult — it's chaotic. A model of predator-prey interaction, modulated by seasonality, simulates with some accuracy the observed three-to-five year population cycles of northern voles. Put another way, the field data can also be broken down by nonlinear analysis to reveal their underlying chaotic components. Rodent population dynamics, whether predicted or observed, are chaotic, but with a strong periodic component imposed by the density-dependent appetites of their mustelid tormentors. Interestingly, chaos is absent in vole populations further south, where the climate is more clement and the predators more varied. H.G.

Guariglia *et al.* which yields the opposite result to theirs. But observations by Mesulam<sup>4</sup> and Anderson<sup>5</sup> suggest that the reverse result may yet be found.

A traditional test for neglect requires the patient to draw a clockface from memory. The figure on page 193 indicates each of the two results typically found in neglect: the patient either leaves out the numbers on the left, or transposes them to the right side of the clockface. The examples given (from one of our own patients) show omission in the acute phase after stroke, followed by transposition in the chronic phase.

Anderson<sup>5</sup> has reported a patient with stroke-induced damage of the right parietal lobe and right thalamus who reliably produced clock-drawings of the second (transposed) type when tested, as is customary, with the eyes open. But when required to perform the same task with the eyes closed, her clockface was drawn normally, with all 12 numbers appropriately placed around the full circumference. Anderson argues (metaphorically) that "right-sided external percepts are more sticky than internal images", but the results of Guariglia *et al.* (on a different patient) show the right side of images to be the more "sticky".

In the case of Guariglia *et al.*, the site of brain damage was right fronto-temporal. This too is surprising, as most previous studies of the generation of visual imagery have shown that left occipito-temporal regions are maximally involved in this process. These areas were intact in the patient of Guariglia *et al.* Accordingly, the authors suggest that their patient can generate "the global image of an object", but that this global representation does not suffice to place specific (local) items

correctly within the overall space. Other recent results, using conventional cancellation tests for neglect, are compatible with this conjecture. Patients with neglect can accurately mark the four corners of a sheet of paper that contains small (local) lines pseudorandomly scattered across the full dimensions of the page. Yet immediately afterwards the patient fails to cross out the small lines on the left of the sheet while having no difficulty cancelling those on the right<sup>6</sup>.

This interplay between global and local perception provokes a question that no study of imaginal neglect has yet addressed: what shape does the patient of Guariglia *et al.* imagine the Piazza dei Cinquecento in Rome to be? If global perception and memory are intact, he should draw (or verbally describe) it correctly; if the global image is 'neglected', the piazza should shrink in the lateral dimension. Whatever the result, the evidence is clearly mounting that in neglect, as in the visual agnosias ('seeing without recognizing')<sup>7</sup>, visuospatial perception and visuospatial imagery do not draw upon identical neuronal substrates. □

John C. Marshall and Peter W. Halligan are in the Neuropsychology Unit, University Department of Clinical Neurology, Radcliffe Infirmary, Oxford OX2 6HE, UK.

- Guariglia, C., Padovani, A., Pantano, P. & Pizzamiglio, L. *Nature* **364**, 235–237 (1993).
- Grusser, O.-J. & Landis, T. *Visual Agnosias and Other Disturbances of Visual Perception and Cognition* (Macmillan, London, 1991).
- Bisiach, E. & Luzzatti, C. *Cortex* **14**, 129–133 (1978).
- Mesulam, M.-M. *Principles of Behavioral Neurology* (Davis, Philadelphia, 1985).
- Anderson, B. *Neurology* **43**, 213–216 (1993).
- Halligan, P. W. & Marshall, J. C. *Cortex* **29**, 167–174 (1993).
- Gurd, J. M. & Marshall, J. C. *Nature* **359**, 590–591 (1992).

## DAEDALUS

## Negative energy

A HOAX of the 1960s claimed that some foods, such as hard-boiled eggs, used up more calories in being digested than they provided in food value. So by eating enough of them you could lose weight. Overweight egg-lovers were soon sadly disillusioned; but Daedalus now proposes a food that works this elegant trick in reality.

He points out that digestion is a process of hydrolysis. It breaks down fats, carbohydrates and proteins into their small component molecules simply by reacting them with water. The reaction evolves little energy and is easily reversible, so that the body can later recombine the components into fats, carbohydrates and proteins of its own. DREADCO chemists are devising a fat whose components combine spontaneously. The fatty acid component is sterically designed to enfold the glycerol part; the molecules attract each other and combine rapidly, giving out energy. To hydrolyse them apart, this energy must be supplied again.

Incorporated into cooking oil, chocolate and so on, DREADCO's 'Thinfat'® will transform dieting. The dieter's digestive system will invest the required amount of energy to hydrolyse it to glycerol and free fatty acid. The separated components will then promptly combine again, liberating that energy as heat. The digestive enzymes will set to work to hydrolyse the stuff again, only to have their work promptly undone a second time. The cycle will continue indefinitely. In metabolic terms, Thinfat has a negative calorific value.

A Thinfat user should enjoy his diet. Instead of indigestion, he will have the novel experience of hyperdigestion. His warm and comfortable after-dinner lethargy will persist for hours, as energy is drained from his body and released as heat. It will be a wonderfully lazy way to lose weight.

Even when the digestive system has won its exhausting battle, Thinfat will continue its work. Once inside the body, it will be stored in the normal fat deposits. There it will form a calorific Trojan horse. If the subject attempts a little exercise, these deposits may be called upon to supply energy. But fat must be hydrolysed before it can be released. Once again the endless cycle of hydrolysis and recombination will be set up; once again energy will be lost from the body as heat. The subject will find his exercise warming but not particularly exhausting — exhaustion is a muscular effect. But it will slim him dramatically. Daedalus may have invented a diet and exercise regime that really works!

David Jones



# Horsepower from a horse

**SIR** — Recent studies of flying animals carrying loads<sup>1</sup> and of skeletal muscle *in vitro* subject to cyclic motion<sup>2</sup> suggest that the maximum sustainable mechanical power per kg of muscle is 100–200 W. Given an animal's size and its proportion of muscle mass, it is thus possible to calculate an upper limit to an animal's power output. This led us to wonder how much horsepower one horse can actually produce.

The body mass of horses varies from less than 100 kg for ponies to more than 800 kg for large draught animals. According to Munro<sup>3</sup>, skeletal muscle for a horse is about 45 per cent of the total mass, but we estimate that only 30 per cent could be used for mechanical work at any one moment. Assuming a mass-specific rate of 100 W kg<sup>-1</sup> of muscle and a body mass of 600 kg, one horse could, in theory, produce 18,000 W or, since one horsepower (HP) equals 746 W, about 24 HP! Is it possible that one horse generates that much horsepower? The assumptions, in the worst case, might inflate the result by a factor of 2, yet this still gives an estimate of about 12 HP. This raises the question: was the definition of horsepower based on a lower rate of work, or can a healthy horse actually produce more than 10 HP?

As to the first possibility, it was James Watt himself who defined horsepower. According to Dickinson<sup>4</sup>, in the early 1780s Boulton and Watt were manufacturing rotary steam engines that replaced horse gins. Quite naturally, payment for the engine was an annual premium based on the number of horses needed to do the equivalent amount of work. In discussions with millwrights, Watt learned that during a day's work a horse would walk an average of two and a half times per minute around a 24-ft diameter mill wheel. Dickinson<sup>4</sup> (p. 145) says Watt assumed a horse exerted a tractive effort of 180 pound force (lbf), yielding a power estimate of 33,929 ft-lbf min<sup>-1</sup> (power = force × distance/time). In Watt's blotting and calculation book this number was rounded to 33,000 ft-lbf min<sup>-1</sup>, equivalent to the more familiar definition for HP of 550 ft-lbf s<sup>-1</sup>. (The US Bureau of Standards<sup>5</sup> gives a different account of Watt's calculation

that says he considered engine friction.) By either calculation, Watt's measure of power output is clearly based on a rate that horses could maintain for a full day, not a peak performance.

As to the second possibility, Collins and Caine<sup>6</sup> list data from the horse pulling contest at the 1925 Iowa State Fair showing that peak mechanical power output of a horse is 12–14.9 HP. This effort lasted only a matter of seconds and is probably a realistic estimate of peak performance. Similar maximal rates, when expressed per kg of body mass, have been documented in human athletes<sup>7</sup>.

Why is the daily work rate so much lower? Collins and Caine<sup>6</sup> suggest that a draught horse should pull 10 per cent of its body weight at a rate of 2.5–3 miles h<sup>-1</sup> (10-hour working day) to maintain health

and vigour. Comparable working rates were suggested by Youatt<sup>8</sup> in 1826. Interestingly enough, both these work rates are just about 1 HP. Furthermore, they correspond to a daily metabolic rate of about 4 times the basal rate, which is a rate that has been documented in other vertebrates performing sustained activity<sup>9,10</sup>. In summary, it seems that the millwrights of the 1780s knew how to keep their animals in good shape, that Watt made his estimates carefully, and that a horse can provide significantly more than one horsepower.

**R. D. Stevenson**

Department of Biology,  
University of Massachusetts at Boston,  
Boston,  
Massachusetts 02125-3393, USA

**Richard J. Wassersug**

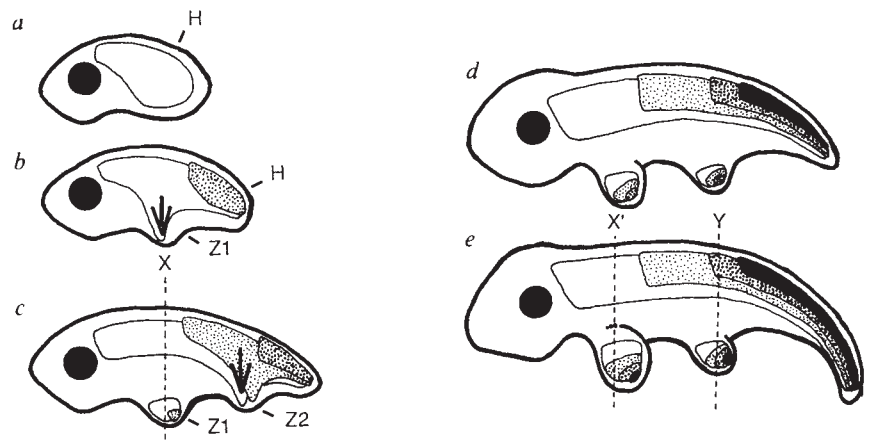
Department of Anatomy & Neurobiology,  
Dalhousie University, Halifax,  
Nova Scotia B3H 4H7, Canada

## Hox genes, fin folds and symmetry

**SIR** — Tabin and Laufer<sup>1</sup> suggest an evolutionary explanation of the similarity between tetrapod fore and hind limbs in which the *Hox* gene-regulated limb pattern originated in the pelvic appendage, and subsequent ectopic activation imposed the regulation and resultant homeotic transformation on the ancestral pectoral fin/fore limb. I agree with the

general proposal that patterns of genetic regulation provide a new level of explanation for homology<sup>2</sup>, but I question Tabin and Laufer's specific evolutionary hypothesis.

The 'pelvic before pectoral limb' evolutionary model is based partly on the continuous lateral fin-fold theory of the origin of paired vertebrate appendages,



Schematic representation of spatial and temporal (a–e) sequential overlapping *Hox* gene expression domains in a developing vertebrate embryo (from refs 2, 8 in ref. 1). Black dot, cranial region; arrows, mesoderm migrating laterally into emergent limb bud; thin outlines delimit *Hox* gene expression domains in limb buds and trunk/caudal regions (shaded from clear, stippled to black). Maintenance and elaboration of *Hox* gene expression along cranio-caudal and limb-outgrowth axes appears to be influenced by a unified system<sup>14</sup> in which signalling activity emerges from centre H, and spreads into flank tissue centres Z1 and Z2 (ref. 14), with peak activity occurring in pectoral before pelvic limb fields. At an early stage (c), *Hox* gene expression in the pectoral limb bud mesoderm resembles that in the axial mesoderm (section X) because the former is derived from the latter. At a later state (e) increased dissimilarity between the expression patterns (section X') results from the fixed anterior position and continued distal development of the pectoral limb bud, and the elaboration of trunk *Hox* gene expression in more posterior axial mesoderm. Initiation of the pelvic limb bud at a later stage incorporates posterior axial mesoderm in which a fuller complement of *Hox* genes has either already been or will be expressed. Position in the posterior flank region leads to closer similarity of expression pattern with adjacent mesoderm (section Y). As the activity of Z1 peaks before Z2 during ontogeny, a similar sequence is argued to have occurred in phylogeny. Contrary to the suggestion in ref. 1, no homeotic transformations, translations of signalling centres between fins, taxa with continuous lateral fin folds or with pelvic but without pectoral fins<sup>3</sup>, are required. However, differences remain between the *Hox* expression patterns predicted by this alternative hypothesis and those few which are actually known. Here, anterior-most genes expressed axially have principal expression in the limb buds, followed by the remainder of the *Hox* network, whereas in known amniote limb buds only the posterior members of clusters A and D of the *Hox* network are expressed.

1. Ellington, C. P. *J. exp. Biol.* **160**, 71–91 (1991).
2. Stevenson, R. D. & Josephson, R. K. *J. exp. Biol.* **149**, 61–78 (1990).
3. Munro, H. N. (ed.) in *Mammalian Protein Metabolism* 133–182 (Academic, New York, 1969).
4. Dickinson, H. W. *James Watt* (David & Charles, Newton Abbot, 1967).
5. US Dept. of Commerce *Cir. Bureau Stand.* **34** (1912).
6. Collins, E. V. & Caine, A. B. *Iowa Agric. exp. Sta. Bull.* **240**, 193–223 (1926).
7. Vandewalle, H., Peres, G., Heller, J., Panel, J. & Monod, H. *Eur. J. appl. Physiol.* **56**, 650–656 (1987).
8. Youatt, W. *The Horse* (Knight & Co, London, 1846).
9. Drent, R. H. & Daan, S. *Ardea* **68**, 225–252 (1980).
10. Peterson, C. C., Nagy, K. A. & Diamond, J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2324–2328 (1990).

and partly on the observation that *Hox* gene expression in the hind limb bud resembles more closely that of the adjacent trunk mesoderm than does that of the fore limb bud<sup>1,3</sup>; Tabin and Laufer seek corroboration from Ahlberg's discussion<sup>4</sup> of coelacanth median (and caudal) fin symmetry, and the earliest record of near-tetrapod limb-like fragments<sup>5</sup>.

The continuous lateral fin-fold theory lacks support from living and fossil organisms, and the use of such hypothetical ancestral archetypes is ultimately incompatible with the continuity of evolutionary change<sup>6</sup>. Apparent lateral fin folds in actual fossil jawless fishes are either restricted to specialized lineages<sup>7</sup> or occur in conjunction with pectoral fins (osteostracans), questioning an ancestor-descendant relationship between the two structures<sup>8</sup>. The numerous inter-girdle spines of climatiid acanthodians (a group of early jawed fishes), long held as evidence of the continuous fin-fold, now appear to be a derived specialization<sup>9,10</sup>.

There is no evidence to support the theory that the pelvic appendage antedates the pectoral<sup>3</sup>. A recent review of agnathan interrelationships<sup>7</sup> suggests that the pectoral fin evolved before any other component of the appendicular skeleton (thus exemplifying the role of fossil data in challenging the validity of decisions concerning the polarity of neontological data<sup>9</sup>). The earliest known pelvic fin skeletons from members of the main vertebrate groups consist of diminutive or abbreviated reiterations of the pectoral pattern<sup>9,11,12</sup>. The strong dissimilarity suggested by Tabin and Laufer is not found.

An alternative hypothesis sketched out in the figure could account for the actual patterns of morphology in the fossil record without recourse to theoretical ancestors, and for patterns of gene expression in embryology without the ectopic initiation of a new pectoral signalling centre. It may also incorporate more easily the phylogenetic history of the fourfold expansion of the *Hox* network occurring between invertebrates and tetrapods<sup>13</sup>. Greater similarity between the expression patterns in the hind limb bud and adjacent axial mesoderm than between the corresponding anterior tissue domains could result

simply from the dynamic sequence of ontogenetic (and phylogenetic) development. Like the general anterior-to-posterior gradient of vertebrate development, the occurrence of pectoral before pelvic appendages could be interpreted as phylogenetic recapitulation, just as the earliest jawed and jawless fishes tend to consist of elaborate brain cases and gill arches with only poorly differentiated axial and tail skeletons.

**Mike Coates**

University Museum of Zoology,  
Downing Street,  
Cambridge CB2 3EJ, UK

**SIR** — Tabin and Laufer<sup>1</sup> present a thought-provoking model for the evolution of the tetrapod limb, but nevertheless leave untackled some important issues.

First, it has long been known that the axial level of the paired appendages of fishes is plastic, in that the specific level of origin of the fins has moved rostrally and/or caudally along the body axis during phylogeny. Therefore, there is no 1:1 correspondence between segment identity and the anterior-posterior point of developmental origin of the fins (see refs 2, 3). If the *Hox* code of the posterior appendage reflects the level of origin of the fin from within the continuous primitive lateral fin fold, then clearly axial *Hox* code and appendage *Hox* code can become dislocated, both phylogenetically and ontogenetically. If the Tabin and Laufer model is correct, then such dislocation is likely to have been a feature of not only the fin-to-limb transition but also of the radiative evolution which has resulted in such diversity of fin phenotype and axial position.

Second, in the three-stage model proposed, the key step is the acquisition of a 'posterior' *Hox* code by the anterior appendage. An explanation for how this might come about is fundamental to the model. It is suggested that the ectopic initiation, in the anterior appendage, of a signalling site such as the zone of polarizing activity, assumed to be formerly confined to the posterior appendage, could provide the means whereby a posterior appendage genetic programme (deploying *Hox D* cluster genes) could be 'co-opted' for the development of the anterior appendage. Indeed, there is now good evidence that some vertebrate homeotic transformations are generated by posteriorization of the relevant *Hox* code, such that inappropriate expression of a posterior code at more anterior levels changes the positional identity of the structures that form at those levels to a posterior character<sup>4,5</sup>. However, the model proposed begs the question of what happens to the intermediate part of the continuous primitive fin fold and fails to address the issue of why the putative homeotic change at anterior levels of the

fold involved acquisition of the most posterior code, rather than the *Hox* code characteristic of the intermediate part of the fin fold. As a corollary, we might ask what type of pectoral appendage did the original anterior *Hox* code specify and can we identify such a phenotype in the fossil record?

Third, any model addressing the problem of the fin-to-limb transition during evolution should acknowledge that mesenchyme in fins and limbs arises from different lineages<sup>6,7</sup>. The mesenchyme of the tetrapod limb is mesoderm-derived and forms the endoskeleton, whereas the mesenchyme of the (phylogenetically more ancient) fin bud is both mesoderm- and neural crest-derived, forming endoskeleton proximally and dermal skeleton distally. In terms of phenotypic evolution of the primitive lateral fin fold, and the fin-to-limb transition, the lineage derivation of the mesenchyme has important implications with regard to mechanisms invoked to explain these evolutionary events. For example, is skeletal pattern within the neural crest-derived (ecto) mesenchyme of the fin specified in a manner identical to the mesoderm-derived mesenchyme? Can we expect to find nested patterns of *Hox A* and *D* cluster genes in the distal mesenchyme of the fin bud comparable to those reported for the distal mesenchyme of the limb bud<sup>8,9</sup>? Although the model proposed by Tabin and Laufer is not incompatible with this embryology, it will need to accommodate the mixed lineage derivation of the mesenchyme found in the appendages of non-tetrapod vertebrates.

To put the Tabin and Laufer model and the foregoing discussion into some perspective, the existence of a primitive lateral fin fold in an ancestral form, although widely discussed, remains entirely hypothetical in the absence of any supporting evidence (palaeontological or biological). Nevertheless, we believe that the authors are correct in stressing the need for comparative studies on *Hox* gene expression using other systems, but urge that questions such as those raised here form an integral part of any such study.

**Peter Thorogood**

**Patrizia Ferretti**

Developmental Biology Unit,  
Institute of Child Health,  
30 Guilford Street,  
London WC1N 1EH, UK

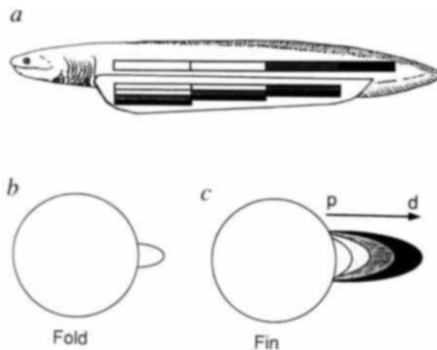
1. Tabin, C. & Laufer, E. *Nature* **361**, 692–693 (1993).
2. Akam, M., Dawson, I. & Tear, G. *Development* (Suppl.) **123**–133 (1988).
3. Tabin, C. *Development* **116**, 289–296 (1992).
4. Ahlberg, P. E. *Nature* **358**, 459 (1992).
5. Ahlberg, P. E. *Nature* **354**, 298–301 (1991).
6. Rowe, T. & Gauthier, J. *Syst. Biol.* **41**, 372–378 (1992).
7. Forey, P. & Janvier, P. *Nature* **361**, 129–134 (1993).
8. Forey, P. *J. Vert. Paleont.* **4**, 330–343 (1984).
9. Maisey, J. G. *Cladistics* **2**, 201–256 (1986).
10. Long, J. A. *Zool. J. Linn. Soc.* **87**, 321–339 (1986).
11. Zangerl, R. *Chondrichthyes I. Handbook of Paleichthyology* (ed. Schultze, H.-P.) **3A** (Fischer, Stuttgart, 1981).
12. Carroll, R. L. *Vertebrate Paleontology and Evolution* (Freeman, New York, 1988).
13. Holland, P. *Bio Essays* **14**, 267–273 (1992).
14. Hornbruch, A. & Wolpert, L. *Development* **111**, 725–731 (1991).

1. Tabin, C. & Laufer, E. *Nature* **361**, 692–693 (1993).
2. Goodrich, E. S. *Studies on the Structure and Development of Vertebrates* (Macmillan, London, 1930).
3. de Beer, G. R. *Homology: an Unsolved Problem* (Oxford University Press, 1971).
4. Kessel, M. & Gruss, P. *Cell* **67**, 89–104 (1991).
5. Lufkin, T. et al. *Nature* **359**, 835–841 (1992).
6. Thorogood, P. in *Developmental Patterning of the Vertebrate Limb* (eds Hinchliffe, J. R. et al.) **347**–354 (Plenum, New York, 1991).
7. Smith, M. M. & Hall, B. K. *Biol. Rev.* **65**, 277–374 (1990).
8. Dolle, P. et al. *Nature* **342**, 767–772 (1989).
9. Yokouchi, Y. et al. *Nature* **353**, 443–445 (1991).



**SIR** — In agreement with Tabin and Laufer<sup>1</sup>, I assume that a fin fold had *Hox* expression domains identical to the parallel *Hox* patterns along the body axis, that is, in a rostral to caudal direction. I would like to suggest an alternative mechanism for the transition from a fold to a new, secondary axis (see figure).

The outgrowth of a simple fold to a more fin-like structure, still extending the length of the animal, must have involved reorientation of overlapping *Hox* expression domains into a proximal to distal direction. I propose that this occurred in response to a growth stimulus exerted by the distal ectoderm of the fold, which may have functionally combined biological elements of the apical ectodermal ridge and the zone of polarizing activity. The *Hox* codes of the respective axial levels of the body axis<sup>2</sup> would have served as the basis



*Hox* codes in folds and fins **a**, Fish with lateral fin along the body length (modified from ref. 1). **b**, Cross-section at an anterior level through a fish with a simple fold and **c**, through a fish with a lateral fin. Anterior *Hox* codes are white; more posterior *Hox* codes grey or black; arrow, proximal (p) to distal (d) axis.

for the extension of codes. *Hox A* and *D* genes typical for the posterior body region thus were activated also at an 'ectopic', relatively rostral level. A broad axis along the side line of a fish could evolve into separate anterior and posterior fins, the precursors in evolution to the fore and hind limbs. Each separate appendage already possessed unidirectional *Hox* gene expression patterns at the time of separation, with expression of the 5' *Hox A* and *D* genes in the most distal area.

One could further speculate that spatial separation of the site of growth stimulation from the centre of *Hox* activation allowed the development of hand or foot plates, that is, the anterior-posterior polarity in the limb axes typical of higher vertebrates.

**Michael Kessel**

Max-Planck-Institut für biophysikalische Chemie,  
Abteilung Molekulare Zellbiologie,  
D-3400 Göttingen, Germany

1. Tabin, C. & Laufer, E. *Nature* **361**, 692–693 (1993).
2. Kessel, M. & Gruss, P. *Cell* **67**, 89–104 (1991).

## Climate change and soil erosion

**SIR** — O'Hara *et al.*<sup>1</sup> ascribe changes in the rate of sedimentation recorded in Lake Pátzcuaro, central Mexico, to soil erosion promoted by cultivation, and use their findings to challenge the common view that prehispanic agriculture was environmentally more benevolent than later versions.

Granted that agricultural and other kinds of land use influenced sediment yield, the possibility remains that shifts in the seasonal distribution and intensity of rainfall had some effect on the depositional narrative. O'Hara *et al.* themselves present evidence for drier conditions between two of their spells of rapid sedimentation. Recent episodes of aggradation have long been recognized in central Mexico<sup>2</sup>. In the Central Plateau many valleys display extensive alluvial fills whose ages are consistent with deposition between AD 500 and 1700 (refs 3, 4). The fills display slit-clay depletion and are better stratified than their source deposits, features more consistent with sustained stream discharges than with accelerated soil erosion. The explanation might then lie primarily in shifts in the relative importance of combined winter and autumn rainfall produced by latitudinal displacements of the subtropical high-pressure cells<sup>5</sup>.

In view of current concerns about climatic instability, it seems surprising that O'Hara *et al.* ignore atmospheric factors when evaluating a uniquely detailed depositional sequence.

**C. Vita-Finzi**

Department of Geological Sciences,  
University College London, Gower Street,  
London WC1E 6BT, UK

1. O'Hara, S. L., Street-Perrott, F. A. & Burt, T. P. *Nature* **362**, 48–50 (1992).
2. De Terra, H., Romero, J. & Stewart, T. D. *Tepexpan Man* (Viking Fund, New York, 1949).
3. Vita-Finzi, C. *Nature* **227**, 596–597 (1979).
4. Vita-Finzi, C. *Geol. Rdsch.* **66**, 99–120 (1977).
5. Wallén, C. C. *Geogr. Annl.* **37**, 52–58 (1955).

**O'HARA ET AL. REPLY** — In our paper<sup>1</sup>, we discussed only the upper 3.36 m (5,000–0 years BP) of the Mastercore. The full 44,000-year-long record from this core shows that environmental changes since the appearance of maize (*Zea mays*), about 3,600 years BP, were more dramatic than those associated with either interstadial/glacial or glacial/interglacial shifts in climate. We note a striking increase in magnetic susceptibility as well as a marked change in the sediment chemistry, with the material accumulating in the lake changing from predominantly authigenic to predominantly allogenic in origin. Land degradation after 3,600 years BP is also reflected in the pollen<sup>2</sup> and

diatom (J. P. Bradbury, unpublished data) records. During the late Holocene, pollen of grassy and weedy species increased to levels even higher than those encountered under the arid conditions of the last glacial maximum (LGM, about 25,000–11,000 years BP)<sup>2</sup>.

Although lacking the spatial resolution of the Lake Pátzcuaro study, cores from other nearby lakes indicate a significant increase in sediment accumulation after 4,000–3,000 years BP<sup>3–6</sup>. The evidence from the crater lake La Piscina de Yuriria, Guanajuato<sup>6,7</sup>, is particularly clear. In parallel with Lake Pátzcuaro, this lake has experienced three phases of accelerated erosion since the advent of maize cultivation. Its diatom and sediment chemistry records, which have century-scale resolution, indicate that in each case accelerated erosion began under wet conditions and terminated after the onset of dry ones. Under natural conditions of closed pine/oak forest with good ground cover, it seems unlikely that an increase in annual rainfall would enhance erosion rates. The opposite is more probable. We have therefore argued<sup>7</sup> that the relationship between climate and sediment yield in this area was modulated by climatically driven fluctuations in the northern frontier of summer-rain-fed agriculture.

Vita-Finzi suggests that changes in the seasonality of precipitation might provide an explanation for increased erosion. Although Lake Pátzcuaro receives very little winter precipitation at present, during the LGM, suppression of the summer monsoon coupled with an increased incidence of *nortes* (cold polar outbreaks) may have resulted in increased winter rainfall. The palaeolimnological record from Lake Pátzcuaro, however, indicates that at this time the lake was clear, acid and unproductive with a low input of sediment from the juniper-sagebrush steppe on the surrounding slopes. There is no evidence from Lake Pátzcuaro, or indeed from any other central Mexican lakes that we have investigated, to suggest that climatic change has had a significant, direct impact on erosion rates.

**Sarah L. O'Hara**

Department of Geography,  
University of Sheffield,  
Sheffield S10 2TN, UK

**F. Alayne Street-Perrott**

**Timothy P. Burt**

School of Geography,  
Mansfield Rd, Oxford OX1 3TB, UK

1. O'Hara, S. L., Street-Perrott, F. A. & Burt, T. P. *Nature* **362**, 48–51 (1993).
2. Watts, W. A. & Bradbury, J. P. *Quat. Res.* **17**, 56–70 (1982).
3. Metcalfe, S. E. *et al. J. Paleolimnol.* **5**, 197–218 (1991).
4. Bradbury, J. P. *Limnol. Oceanogr.* **16**, 180–200 (1971).
5. Metcalfe, S. E. *et al. Geoarchaeology* **4**, 119–141 (1989).
6. Metcalfe, S. E. & O'Hara, S. L. *Ingeniería hidrául. Méx.* **7**, 107–121 (1992).
7. Metcalfe, S. E. *et al. in Effects of Environmental Change in Drylands* (eds Millington, A. C. & Pye, K.) (Wiley, in the press).

## Adsorption at domain edges

**SIR** — We recently presented high resolution scanning force microscope (SFM) images of a phase-separated, thin-film surface composed of circular hydrocarbon domains surrounded by a fluorocarbon sea<sup>1</sup>. We have now used these well-characterized films as substrates for preferential adsorption of macromolecules. As shown in the figure, the macromolecules (tobacco mosaic virus, TMV, in this example) are dispersed and readily visible against the distinct background of the phase-separated Langmuir-Blodgett (LB) film. Adsorption occurs preferentially on the fluorocarbon plane, at the juncture between hydrocarbon and fluorocarbon domains. This definitive assignment of binding loci is made possible by the highly localized resolution of the imaging instruments. We can determine distribution and orientation of the adsorbate.

Close inspection of the conformation of the macromolecule reveals that (1) it typically nestles against the hydrocarbon domains, adhering to the side of the domain, (the island-like domain protrudes about 1.5 nm above the surrounding fluorocarbon sea in this particular film); and (2) it can bend around these hydrocarbon islands. Because TMV is a rigid, rod-like structure, the distortion indicates that the adsorption force is strong.

Whereas it is curious that these biological macromolecules lie preferentially

'on' the fluorocarbon surface, their actual location is the juncture between two domains, that is, at a step site. An increase in substrate-adsorbate contact area results from adsorption at a step, a probable contribution to the choice of these sites in the adsorption process.

Another likely source for this selectivity is the ordering of surface energies of the available sites for adsorption. One can regard these phase-separated surfaces as being of three kinds: those terminating in  $-\text{CF}_3$ ;  $-\text{CH}_3$  and  $-\text{CH}_2-$ . Most of the film surface is accounted for by the first two, comprising the terminal methyl groups of the respective carboxylates that comprise the LB film. The third surface is the exposed sides of the hydrocarbon domains, composed of methylene groups. Comparison of the surface energies of the three surfaces indicates that the methylene-terminated surface should present the highest energy surface of the three: the critical surface tensions are 6, 22 and 31 dyn  $\text{cm}^{-1}$  for  $-\text{CF}_3$ ,  $-\text{CH}_3$ , and  $-\text{CH}_2-$  terminated surfaces, respectively<sup>2</sup>.

The selective adsorption of the macromolecules against the hydrocarbon domains could be explained on the basis of the highest surface energy site available. The selectivity for adsorption at the 'side-walls' has been observed in our laboratory with other adsorbates as well — DNA strands, for example, adsorb in extended conformations. Control experiments of TMV adsorption onto mica, glass, graphite and fully fluorinated surfaces

have produced SFM images dominated by aggregates and rigid-rod conformations.

The image pictured here is obtained in a non-contact imaging mode in which the probe rides about 1–2 nm above the sample surface, avoiding physical contact between probe and sample (ref. 3, and D.A., manuscript submitted). Therefore, the possibility of physical 'placement' of the adsorbate on surface sites by the scanning action of the probe can be eliminated.

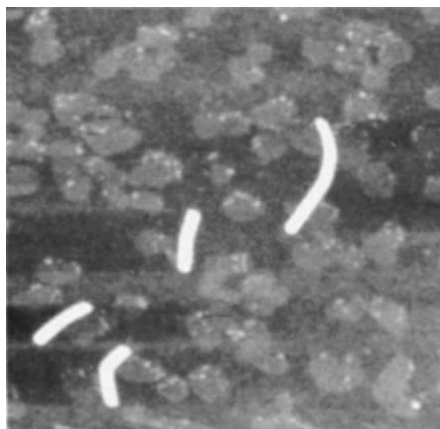
This substrate/adsorbate system serves as an example of a crafted surface of heterogeneous yet carefully controlled molecular properties. More generally, the surface could be tailored by the choice of its molecular components to result in well-defined gradients in surface energy, and hence a selective and patterned adsorption.

**J. Frommer, R. Lüthi, E. Meyer, D. Anselmetti, M. Dreier, R. Overney, H.-J. Güntherodt**

*Physics Institute, University of Basel, Klingelbergstr. 82, CH-4056 Basel, Switzerland*

**M. Fujihira**  
*Department of Biomolecular Engineering, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 227, Japan*

1. Overney, R. *et al.* *Nature* **359**, 133–135 (1992).
2. Zisman, W. A. *Adhesion and Cohesion* (ed. Weiss, P.) 186 (Elsevier, New York, 1962).
3. Martin, Y., Williams, C. C. & Wickramasinghe, H. K. *J. appl. Phys.* **61**, 4723–4729 (1987).



SFM image ( $1.1 \times 1.1 \mu\text{m}^2$ ) of TMV adsorbed on to the surface of a mixed hydrocarbon/fluorocarbon LB film. In the background, the lighter, rounded features are the hydrocarbon domains of the phase-separated film; their darker surroundings are fluorocarbon phase<sup>1</sup>. The bright, elongated features are the TMVs; the diameter of their rod-like geometry is about 18 nm (though SFM tip effects widen this dimension somewhat) (D.A., manuscript submitted). Sample prepared by evaporation of a dilute aqueous solution under ambient conditions.

## Errors and ozone measurement

**SIR** — Mimms<sup>1</sup> recently compared ozone measured using his hand-held TOPS instrument with ozone data from the Total Ozone Mapping Spectrometer (TOMS) on Nimbus 7 and noted possible errors in the TOMS data. The TOMS ozone data used in Mimms's analysis were preliminary and should not have been used for quantitative analysis.

In 1992 the Nimbus 7 orbit had drifted such that we could not measure direct sunlight at the satellite, a measurement needed to maintain the calibration of TOMS to high accuracy. Consequently, between 12 February, when we lost the Sun, and 30 September, when we re-acquired it, we were using an extrapolated calibration. These preliminary data are made generally available with a warning that the instrument calibration is not final.

Mimms's analysis in relation to his TOPS measurements showed that TOMS appeared to have as much as +3.7% ozone error. When the final calibration was established for all of 1992 and the final data were processed, we found that the preliminary data had indeed been in error by about +1.7%. The remainder of the

discrepancy appears to have been caused by the effect of Pinatubo aerosols on the TOPS measurement. We recently described the calibration and production of the final TOMS data<sup>2</sup>, and confirmed the final calibration of TOMS, through comparison with the world standard Dobson instrument (183).

Mimms's comparison of his TOPS measurement with TOMS provided a valuable early indication that the extrapolated TOMS calibration was in error. The satellite ozone measurement programme will always need ground measurements to monitor the calibration of the satellite instruments. But we must emphasize that the 'real-time' data we produce and make generally available should be used with caution and that only 'final' data should be used for quantitative ozone analysis.

**Richard D. McPeters**  
**James F. Gleason**  
*NASA/Goddard Space Flight Center, Mail Code 916, Greenbelt, Maryland 20771, USA*

1. Mimms, F. M. *Nature* **361**, 505 (1993).
2. Gleason, J. F. *et al.* *Science* **260**, 523–526 (1993).



# Ambitious intentions

Igor Aleksander

**AI: The Tumultuous History of the Search for Artificial Intelligence.** By Daniel Crevier.  
Basic Books: 1993. Pp. 386. \$27.50.

ARTIFICIAL intelligence is unique in the history of science and technology in at least one important way. Its practitioners are entirely absorbed in discovering what can and cannot be done with a complex human invention — the digital computer. This differs from the concerns of, say, an aeronautical engineer or a television set designer in the sense that AI pioneers had set out to produce a *doppelgänger* for biological humans with the simplest of notions of how this could be achieved. Publicly, they made extraordinary claims that to make a machine with the intelligence of a human was merely a matter of making computers bigger and faster. The discovery that their results seemed to fall short of their claims tended to be a much more private affair. As an idle thought, where would an aircraft designer be with a similar record of failure? The answer is that he would be unemployed if not in jail. And yet, even recognizing these errors of judgement, Daniel Crevier in his readable account of AI finds that those who developed the technology over the past 40 years were imaginative and sincere in their beliefs. So how is it that they could be so wrong in their predictions? What purpose did their efforts serve?

These are the questions that Crevier attempts to answer in his speculative history. He draws with good effect on recent interviews with AI gurus such as Marvin Minsky, John McCarthy and Herb Simon. As a former pupil of some of these AI masters, he writes with a good mix of awe for their achievements and scepticism of their pronouncements. This familiarity has both good and bad effects. It gives the accounts a sense of enthusiasm and authority but does not allow Crevier to stand sufficiently far back to remain immune to the seductive idea that the truly intelligent machine is just around the next corner. This leads to contradictions. For example, he writes that in the 1960s "AI blossomed a thousand flowers", only to revise this later as: "At the end of the 1970s the intelligent machine seemed no closer than in the 1960s". If the aims of AI remained elusive, what, then, were the "thousand flowers"?

Crevier suggests that AI led to a deeper understanding of both logical and informal representations of knowledge on computers. The ultimate product was the expert system now commonly used to capture procedures (for example, the configuration of large computers) normally done by humans. But expert systems

depend only on a small part of the effort that went into AI. Crevier further suggests that what is now known as "object oriented programming" (letting parts of a program behave like an object with well-defined properties) had its roots in AI. This is undoubtedly true, as is the fact that interesting computer languages such as LISP and PROLOG are products of AI. But it all has the feeling of spin-off rather than pivotal science. For my part, I agree with another of Crevier's suggestions that one of the main effects of AI has been to highlight the informal but speedy way in which intelligent humans behave and the contrast between this behaviour and what can be done by programming a computer. But is this discovery worth the price of unfulfilled AI promises?

Crevier still believes that, despite the failures so far, greater computer power and investment in research will lead to steady progress towards machines that are smart in the human way. He attempts to argue his case by using computer chess as an example. Chess programs have plodded their way steadily up the US professional ranking from being below average in the early 1970s to now being at master level. The progression seems linear and dependent mainly on brute computer power. This leads Crevier to conclude that in the next 10 years not only will the progression lead to a computer program being the world chess champion, but also that AI will evolve successfully in other areas of endeavour that are said to require intelligence.

Both these conclusions suggest that the author has been seduced by the old AI magic. The day that an AI program beats the world chess champion will make the computer about as interesting as the man on a bicycle at a running race. The computer's mechanical nature will by itself disqualify it from the competition. Crevier's second error is that chess is a game that is so constrained that it requires a very special form of intelligence. Away from the chess board the chess master is no more able to cope with the problems of life than the amateur. So while chess computers win their victories, the machinery may not be able to make common-sense decisions.

Crevier suggests that there are already projects that seek to meet the last criticism. He cites Douglas Lenat's "Cyc" (for encyclopaedia) endeavour. This is a \$25 million program that encompasses 200 person-years of effort. Lenat is taking the

brute-force approach in trying to store a colossal number of basic tenets of common sense — things such as: "People come in two kinds, men and women". Undoubtedly, the bigger and the faster the computer, the more success there will be in accessing and using this knowledge. But Lenat's entire approach may be flawed as it requires storage space roughly in proportion to the number of facts. Recent work in connectionism has shown that distributing knowledge over interacting elements (artificial neurons) could lead to a storage requirement that grows only with the logarithm of the number of facts that need to be stored. It seems likely that the brain would have evolved to use this neat trick. There is no reason why either Lenat or Crevier should be aware of these developments as connectionism has largely been seen as a competing and alien technology by the AI gurus. My own feeling is that the smart machine will have to draw on whatever technology is around and partisan attitudes towards AI will merely lead to systems that are less competent than they might be.

But these are minor quibbles. Crevier's story is a worthwhile tribute to the efforts of those who used their skills and understanding of computing in developing a debate that has not only been technological in nature but has had a considerable influence on philosophy and psychology. In the closing chapters, Crevier asks the interesting question of the pioneers: what do *they* think they were doing? The answer given by Gerald Sussman, the designer in the 1970s of PLANNER, a celebrated robot planning program, shows that there is still much idealism in these retrospective views. He suggests that future historians of science will see AI to have been as great an event as the discovery of calculus: "We are witnessing a breakthrough in the way that people express complex ideas . . . I believe that this new capacity will have a profound influence on humanity over a long period of time and will be the thing that's remembered many years from now." Unfortunately, the examples he gives for handling complexity are the design of electronic circuits and finding the square root of any given number.

Humanity is surely looking for solutions to deeper complex problems such as wars, economic systems and threats to health. The efforts of the AI geniuses of the last 40 years may well prove to be much less significant than Sussman suggests. Nevertheless, this should not prevent anyone from reading about them in Crevier's book and allow themselves to be beguiled by the excitement of a dream. □

*Igor Aleksander is in the Department of Electrical Engineering, Imperial College of Science, Technology and Medicine, Exhibition Road, London SW7 2BT, UK.*

# The early Earth

Euan Nisbet

**Proterozoic Crustal Evolution.** Edited by K. C. Condie. Elsevier: 1993. Pp. 537. DFL280, \$175.

TOWARDS the end of the summer, not long before Labor Day, a curious ritual is to be seen in many geology libraries: the spectacle of flustered Phanerozoic stratigraphers, their suntans melanomically glowing after a summer's dinosaur digging, being overwhelmed by toppling piles of Precambrian journals. It is time to write the early lectures for Earth History 301, those quickly passed-over synopses of the first 4 billion years before the really interesting stuff begins in the Burgess shale. In particular, the Proterozoic eon (from roughly  $2.5 \times 10^9$  to  $0.5 \times 10^9$  years ago) is the dark age of stratigraphy, a nightmare for many of those lecturing in Earth evolution. For those poor teachers, Kent Condie has edited a welcome time-saver — a volume that provides a good general review of the problems of the Proterozoic. For the brave souls who actually specialize in Proterozoic evolution, the volume provides some excellent summaries of active areas of research, and should provoke both thought and interest.

*Proterozoic Crustal Evolution* contains 13 chapters, each tackling a major aspect of the problem. Most of the chapters are written by well-known workers in the field. The volume is the outgrowth of the International Geological Correlation Program Project 217 on Proterozoic geochemistry, so its bias is towards crustal geochemistry, but the spread is wide. There are chapters on the ophiolite problem by H. H. Helmstaedt and D. J. Scott, mafic dykes by J. Tarney, the Bushveld complex and other layered bodies by G. von Grunewaldt and R. E. Harmer, granulites by S. L. Harley, iron formations by C. Klein and N. J. Beukes, and isotopes by P. J. Patchett. Each of these topics has its controversy, and the issues are important not simply for the narrow sphere of Proterozoic geochemistry, but more broadly in the longer sweep of Earth evolution. Did early Proterozoic tectonics emplace ophiolites? Perhaps, but the answer depends in part on how far the definition is stretched. How did the gigantic Bushveld intrusion occur? Is it an ambitious boninite or a perverted komatiite, or some other beast? Are the Proterozoic granulites the bones of ancient 'Tibets'? What moved the iron in ironstones? Why were the anorthosites and anorogenic granites emplaced?

These are only a few of the topics discussed. Other chapters include work on collisional and accretionary orogens, volcanics, xenoliths, rifts and accretion. We

still know far too little about the great length of Proterozoic time. In a semi-familiar world, it is not clear whether strict actualism applies — were Proterozoic orogens strictly comparable to modern parallels? — or whether uniformitarianism applies in the modified sense of a planet evolving within its own internal constraints. Was the temperature of the asthenosphere different, and if so, by how much, and how much thicker did this make oceanic crust? How large and how thick were the rafts of continental lithosphere? To what extent did life control the chemistry of the seas?

Recently, another large volume on the Proterozoic has been published (*The Proterozoic Biosphere* edited by J. W. Schopf and C. Klein, Cambridge University Press, 1992; for a review see *Nature* 361,

601; 1993). The two books will make excellent companions, one concentrating on the biological side of the history of the eon, the other mainly on abiotic processes. The Proterozoic evolution of North America has also been well chronicled by Paul Hoffman in a set of essays for the Decade of North American Geology series and elsewhere. As a general introduction to the Proterozoic, this outpouring should keep any worried lecturer supplied with enough material, and may save her from perishing under a toppling pile of journals. This may even be the year to expand the Proterozoic section of the course. □

Euan Nisbet is in the Department of Geology, Royal Holloway, University of London, Egham, Surrey TW20 OEX, UK.

## The constraints of history

Hugh Freeman

**The Most Solitary of Afflictions: Madness and Society in Britain, 1700–1900.** By Andrew Scull. Yale University Press: 1993. Pp. 448. \$45, £29.95.

"MADNESS", writes Andrew Scull, "seems to attract more than its share of myths"; how unfortunate, then, that his latest work should most egregiously add to them. An Englishman who is now professor of sociology at San Diego University, Scull has devoted virtually his whole career to studying the management of mental illness — in particular, English

nineteenth-century asylums. He burst on the scene in 1977 with *Decarceration*, a short but vitriolic work in which he interpreted the historical rise and fall of the mental hospital primarily in terms of social control. Its ideological underpinning, reflecting the sociology of the time, was strongly Marxist, an aspect of the discipline which has since quietly faded. Two years later, Scull published *Museums of Madness*, a history of the Victorian asylum, of which the present work is an updated edition.

Since then, the author has remained highly productive, contributing to and editing books on this theme; his scholastic energy is prodigious and every piece of his writing is profusely referenced from an immense range of sources. In, for instance, an earlier biographical chapter on John Conolly, the reforming asylum superintendent of the 1840s, Scull greatly advanced knowledge of an important figure who until then had been poorly researched. Yet, regrettably, this work has often seemed to be less of a genuine search for greater understanding of the historical processes of psychiatry than an attempt to gain support for an ideological position taken up *a priori*.

Like German historians embroiled in their *historischstreit* over the interpretation of Nazism, Scull has pursued a vendetta against those — notably Kathleen Jones and Gerald Grob — who did the fundamental research on British and American asylums, respectively. Their scholastic crimes, in his eyes, are first to regard the motives of lunacy reformers as at least significantly humanitarian, and second to portray things as becoming generally much better for the mentally ill in the 250 years from the early eighteenth century. This is dismissed here as "naive

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

Insane treatment — early nineteenth century engraving of Bedlam.

Mary Evans/Society for Psychological Research



Whiggism", yet Scull himself depends repeatedly on Jones as a source, and indeed before her pioneering work of 1972 (*History of the Mental Health Services*, Routledge) — soon to appear in a new edition — the whole subject was largely *terra incognita*. Whether he recognizes it or not, he stands on the scholarly shoulders of those who came before him.

The other main strand in his argument is that from its earliest beginnings, psychiatry has been little more than a gigantic confidence trick, "persuading a wider public to take seriously its claims to possess an expertise resting on a scientific basis". Part of this process, Scull believes, was the establishment of a monopoly of the superintendent posts of asylums in the early 1800s, although his account does not reach the point where doctors were themselves replaced by lay administrators. Even the psychiatry of today is described as "seeking to assure us of the benevolence and rationality of its interventions even as it attempts to lay claim to an ever-widening sphere of social life". The extreme bias of this interpretation, with its unrelieved sarcasm and invective, is reminiscent of *Pravda* in its worst days.

Behind the brilliant verbal pyrotechnics and exhaustive citing of archival sources is a deeply flawed concept. Scull's view is essentially ahistorical. He condemns Grob for pointing out that asylums were little different from most human institutions in their imperfect functioning, on the grounds that these were "the product of historically specific and closely inter-related changes in that society's political, economic and social structure", as well as of intellectual and cultural shifts.

But that is precisely the point. To understand this story, we need to ask: what was the 'success' of any kind of hospital at that time, and were asylums any different? How did the proportions of patients in asylums who became better or, worse, remained unchanged or died, compare with the outcome of those sufferers from mental illness who remained at home, in workhouses or wandering abroad? To what extent were those admitted to asylums as physically ill and neglected as they were mentally disturbed? Were psychiatrists any different from other contemporary doctors in their clinical approach or scientific knowledge, and were they any less effective? Most importantly, how does the evolution of the asylum, with the increasing chronicity of its patients in the nineteenth century, relate to the natural histories of the disorders it treated? These questions are never asked, let alone answered. □

Hugh Freeman is editor of the British Journal of Psychiatry, 17 Belgrave Square, London SW1X 8PG, UK.

## Prescription or poison?

Beverly E. Griffin

**A Dancing Matrix: Voyages Along the Viral Frontier.** By Robin Marantz Henig. Knopf: 1993. Pp. 269. \$23.

THIS book is about viruses that may have existed centuries ago and are now gone, some that emerged more recently and are still with us, and some that will probably yet emerge and need to be faced. In the field of virology, there are few miracle cures and fewer still short, sharp shocks that can be administered to counter an agent that cannot be seen (except with the aid of the electron microscope) or even propagated outside a living cell. So it is sometimes difficult to generate optimism about viruses and public health. Yet Robin Marantz Henig ends her book with a section that not only reads splendidly but also illustrates how a community's insight and creativity can be effective in "fighting back", that is, in predicting a viral outbreak before it occurs or, as she calls it, "anticipating the next AIDS".

She discusses, for example, how chickens are being used in Florida as sentinels to anticipate outbreaks of human viral encephalitis that could result from infection with some mosquito-borne viruses. Because the chickens can also be infected by these encephalitic viruses, they can act as forecasters, alerting physicians to the possibility of the disease appearing in patients and public health authorities of the need to institute an anti-mosquito spraying campaign. Similar viral "listening posts", in an age of rapid transportation of goods and people, established on an international basis, could also be used to anticipate problems that might arise in the transportation of viruses from an "immunologically seasoned" population in one part of the world to a virologically naive population in another. To a limited extent, this approach is now being adapted for the influenza virus. The time may yet come when broader global epidemiologic surveillance is given due consideration.

The harnessing of viruses for delivering 'good' genes to replace 'bad' or faulty ones by gene therapy is also cause for optimism. Henig calls it "viral domestication". Why not? Successful virus domestication probably requires, however, a depth of scientific knowledge that is at present available for only few, if any, viruses. For sure, viruses may be harnessed with only limited knowledge for the advantage of the host, but this would seem to require a kind of quirky luck that has not been a hallmark of the field. So Henig concludes her book by arguing for the creation of a

"new biology" to deal with viruses — one where individual and now more strictly defined disciplines, such as molecular genetics, classical microbiology, immunology and population biology, are amalgamated in efforts to outface this wily microorganism, loved by none (except perhaps genuine card-carrying virologists) but crying out for understanding. To create an institute for such a new biology would, however, require an element of imagination few governments seem to possess.

The title of Henig's book is attributed to Lewis Thomas's *The Lives of a Cell: Notes of a Biology Watcher* (1974). The dancing matrix comes from Thomas's idea that viruses "dart, rather like bees, from organism to organism, from plant to insect to mammal to me and back again" and as such, may be nature's efficient mechanism for keeping a gene pool varied and fluid, stirring up the genetic arrangements of species that might otherwise be staid, inflexible and unable to adapt to changing environments. Evolutionarily, viruses may act for the good of the species. This is not our short-term experience, however; even the word for this microorganism is derived from the Latin for 'poison'.

A virus is the ultimate parasite, totally dependent on its host for existence, and therefore having a vested interest in the survival of that host. Alternatively, it must be sufficiently infectious that it can find a new host before the old one dies. Most of the book therefore concerns the fragile balance between this microbe and man. Henig's excellent chapter entitled "A Virus Primer" introduces this subject, and even the (apparently) complex concept of DNA, RNA and the genetic code, in such a lucid manner that it should be intelligible to all but the most recalcitrant reader. Although trained as a journalist, Henig reveals herself as one who has become deeply and intensely involved in her subject, even to the extent of (almost) becoming anthropomorphic about it: "Viruses are so weird and wily that they fairly cry out for metaphors when you write about them". (This eventually happens to all of us in this field!) There are many examples here of the occasions when the fragile balance between the host and the organism has been upset. They range from the stories of mad cows (maybe a virus) to dead dolphins (definitely a virus), from AIDS to Ebola, and make excellent reading, for this is *not* science fiction. The recent outbreak of an Ebola-like virus among research monkeys in an animal colony in Virginia with its frightening overtones is also discussed here; this outbreak was contained, but only just.

If the book has a central overriding theme, it is about new viruses that may be on the horizon, about 'why' and 'how' they could emerge, and the effect they might have on our own society if we are not prepared to deal with them. Viruses

such as dengue fever, Ebola, Hanta, Lassa — the really nasties — are presented in this context. But this is neither an emotional nor a hysterical book. Henig does, however, emphasize that we, globally, are at the mercy of at least two major events, one at present totally outside our control, the other hopefully not. The first of these, which plays havoc with vaccine design and other therapies, is viral mutation. Viruses, particularly those with an RNA genome, can undergo fairly rapid change (mutation) in response to host or environmental alterations. It may turn out that these mutations are not random and will therefore be within our control, but so far we don't know the rules.

The second event is that created by abrupt changes in the environment — deforestation at the behest of industrial development, virus-carrying agents transported rapidly from one part of the world to another, and so on. Such events bring

rodents and arthropods, both excellent carriers of viruses, into contact with humans in a way that might disseminate disease. Henig discusses, among other examples, the case of the tiger mosquito that was imported into Texas as late as 1985, in a Japanese shipment of used automobile tyres, and has since spread to other states. This mosquito can act as a vector for dengue haemorrhagic fever or yellow fever. Viruses that might not naturally cross species barriers easily may do so if introduced by the bite of a rodent or mosquito. Again, we don't know the rules.

This is a vital and stimulating book. Once picked up, it is difficult to put down — for my money, more gripping than *Jurassic Park*. □

*Beverly E. Griffin is in the Department of Virology, Royal Postgraduate Medical School, London W12 0NN, UK.*

## Art, society and archaeology

*Marc Bermann*

**Chavín and the Origins of Andean Civilization.** By Richard Burger. *Thames and Hudson*: 1992. Pp. 248. £38, \$49.95.

THE period 900–200 BC was one of dramatic change among civilizations of the central Andes. New technologies and political forms were accompanied by the spread of the Chavín style of architecture and sculpture, named after Chavín de Huántar, the extraordinary archaeological site in the northern highlands of Peru where it received its highest degree of elaboration. This distinctive art style wove together stylized jaguars, eagles, serpents, plants and anthropomorphic deities such as the Staff God. Richard Burger now provides an authoritative treatment of the development and significance of this art style.

Burger places Chavín in a broad context. He gives an overview of Andean prehistory from 1800 to 200 BC, describing the larger monuments at other sites and their associated art styles, and reviewing what is known about the societies that built them. His tracing of these antecedents is generally comprehensive, although his view that Chavín civilization represents a cultural transformation occasionally leads to overly broad characterizations. And some archaeologists would disagree with his belief that pre-Chavín populations lacked complex sociopolitical organization or powerful élites. In fact, there has so far been little investigation into the sociopolitical organization of societies during the Preceramic and Initial periods (that is, from 2500–900 BC).

Burger's discussions of the develop-

ment of Chavín de Huántar and of the relationships between the Chavín art style, cosmology and ritual are excellent. Among the highlights are his analysis of previous hypotheses about the coastal or tropical rainforest sources of Chavín iconography, his suggestion that a series



**Headstone — one stage of a shaman's drug-induced transformation.**

of carved stone heads at the site depicts a shaman's drug-induced transformation from human to animal spirit, and the summary of his own research into residential areas at the site. His argument that the Janabarriu occupation in 400–200 BC marked the beginnings of a class system is very interesting, but it rests on small numbers of archaeological samples.

There has been much debate about the spread of the Chavín art style. Burger argues that the style is the "symbolic

expression of a religious ideology", which buttresses the claim that its spread reflects religious rather than political processes. He links the growth of the Chavín cult to changes in the interregional exchange system, the collapse of the pre-Chavín coastal societies of Peru, the growth of economic prosperity in the highlands and the development of political complexity as local élites used the cult to justify their privileges. But some might object to his conclusion that Chavín was the forerunner of later attempts to "forge a single Andean civilization" or that the appearance of the art style marked the emergence of complex societies in the Andes.

Nevertheless, this is a challenging and generally persuasive study. Burger's argument, that religious ideology is not merely an epiphenomenon but may have an important role in shaping society and culture, is one that few archaeologists have addressed with such insight and documentation. Although the volume is remarkably free of jargon, a glossary of terms such as 'complex society' and 'stratified society', would have been useful for the nonspecialist. Burger occasionally uses these terms in a loose, confusing way that fails to indicate the relationships between inequalitarian social organization and differences in wealth and access to resources in state and pre-state societies.

Burger's definition of 'civilization' as "a society with a high level of cultural achievement in the arts and sciences as made visible in the form of material objects" underlines the art-historical orientation of the volume. Yet despite the focus on elaborate iconography and public architecture, Burger successfully discusses Chavín civilization in relation to many broader issues, including the development of social inequality, camelid domestication, metallurgy and shifts in settlement and adaptation.

His use of the term 'civilization' implies a unity that may not have existed. In truth, we cannot really talk about a 'Chavín society' or 'Chavín people' extending much beyond Chavín de Huántar itself. As the book itself demonstrates, one of the more interesting aspects of the Chavín phenomena is the extent to which many different peoples or societies adopted Chavín-style iconography and materials.

The impressive scholarship and interpretive sweep of this volume will ensure its position as the definitive study of Chavín and the evolution of complex societies in the pre-Hispanic Andes. It will also serve as a valuable textbook. Moreover, its appealing writing style, logical organization and excellent illustrations make it an outstanding introduction to early Andean prehistory for the general reader. □

*Marc Bermann is in the Department of Anthropology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.*



# Climate instability during the last interglacial period recorded in the GRIP ice core

Greenland Ice-core Project (GRIP) Members\*

**Isotope and chemical analyses of the GRIP ice core from Summit, central Greenland, reveal that climate in Greenland during the last interglacial period was characterized by a series of severe cold periods, which began extremely rapidly and lasted from decades to centuries. As the last interglacial seems to have been slightly warmer than the present one, its unstable climate raises questions about the effects of future global warming.**

THE Eemian interglacial period is a prime target for palaeoclimate reconstruction<sup>1</sup>. A variety of evidence indicates that during isotope stage 5e of the marine oxygen isotope record (~125 to 115 kyr BP), usually correlated with the Eemian interglacial period in Europe and Sangamon in North America, conditions were at least as warm as today<sup>2,3</sup> and probably fairly stable<sup>4</sup>. Some evidence suggests that there may have been strong regional differences, with cyclical climate changes in areas bordering the north Atlantic<sup>3</sup>. Thus there is some hope of examining the behaviour of the physical and chemical climate of the Earth under conditions rather warmer than present, a possible analogue for future climate evolution.

Polar ice cores drilled in optimum locations allow the reconstruction of highly resolved climate records. The Vostok core<sup>5</sup> from central Antarctica was the first to yield an environmental record spanning the entire last glacial–interglacial cycle. At this site, ice deposited during the Eemian period is still ~300 m thick, and as it lies about halfway through the total ice depth the record can be considered robust against distortion due to flow irregularities close to bedrock. Generally, higher annual snowfall and thinner ice cover in Greenland ensure that the Eemian ice will be found relatively deeper in this ice sheet. The deep Camp Century core from northwest Greenland<sup>6</sup> may have reached the Eemian period, but probably did not penetrate it; in this period, at least the detailed record seems to have been strongly modified by rapid shear in the basal ice.

Here we present data obtained from an ice core drilled close to bedrock at Summit, central Greenland, which penetrates the Eemian period. The data cast doubt on some of our present conceptions of this period. Although the timing and amplitude of low-frequency variations (>5 kyr period) in primary indices of climate and atmospheric circulation throughout the last glacial cycle closely mirror those recorded in the Vostok core, there are striking differences at higher frequencies<sup>7</sup>. These are most marked in the Eemian period, for which the record is highly structured. Most constituents examined so far show evidence for a series of shifts from levels typical of warm interglacial conditions to levels more typical of the mid-glacial period. The changes can be transient (lasting only a few decades to centuries), seemingly analogous to the series of climate 'mode-switches' previously identified in the late glacial period<sup>8</sup>, or they can remain latched for up to 5 kyr.

## Record of the Eemian at Summit

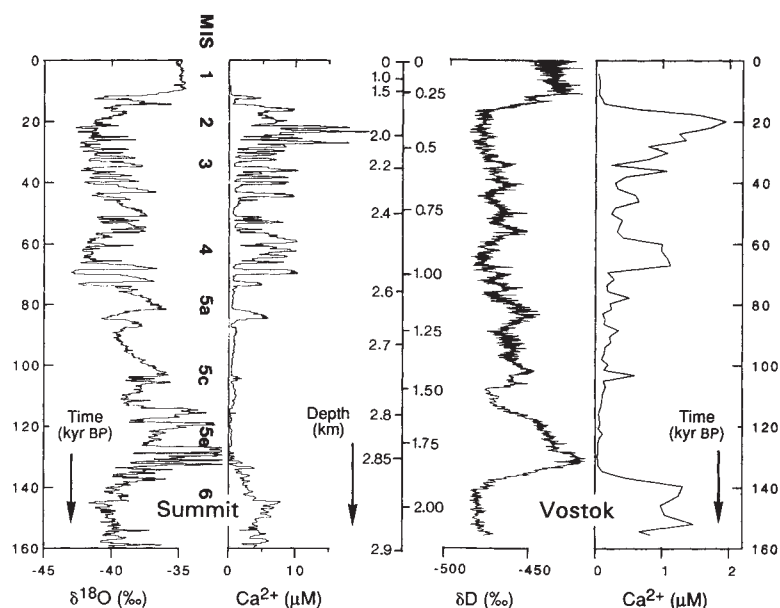
The GRIP ice core, drilled into silty ice close to bedrock at Summit on the main ice divide (72°34' N, 37°37' W) in summer 1992 (refs 7, 9) and the parallel core being drilled about 30 km to the west by the United States Greenland Ice-Sheet Project (GISP2)<sup>10</sup> were positioned to obtain records of the best quality possible through at least the last complete glacial cycle. The stable isotope profile for the complete core is presented in ref. 7. Although detailed analysis of the core is still under way, several constituents were measured continuously at the drill site, or shortly afterwards. These have allowed a partial characterization of the Eemian sequence, which is ~80 m thick and extends from about 163 to 238 m above bedrock. The observed thickness is broadly consistent with model predictions based on the assumption that Greenland was covered by an ice sheet of similar maximum thickness to the present throughout the Eemian interglacial period<sup>11</sup>.

A preliminary chronology for the GRIP core going back to 14.5 kyr BP has been proposed<sup>9</sup> based on the stratigraphy. The chronology has been extended through the Eemian section of the core by ice-flow modelling<sup>7</sup>, and confirmed by demonstrating a close correlation between low-frequency features (>5 kyr period) in the GRIP stable isotope profile with corresponding features in the Vostok profile<sup>5</sup>, in the SPECMAP marine sediment standard isotope curve<sup>12</sup> and in the Devil's Hole terrestrial isotope record<sup>13</sup>. The stable isotope composition is a well established climate indicator, depending to first order on air temperature<sup>5,14</sup>, but also contains information on conditions in the source regions for atmospheric moisture, and hence is linked also to atmospheric circulation changes<sup>15</sup>. However, the closely paralleled behaviour of stable isotopes in Greenland and Antarctica applies to other constituents such as calcium, as can be seen in Fig. 1. Calcium is mainly derived from terrestrial dusts, correlating closely with aluminium<sup>16</sup>, and temporal variations are connected with large-scale changes in the transport and loading of terrestrial dusts<sup>17</sup>. Throughout marine isotope stage 5 (MIS-5) with the exception of a brief period between MIS-5a and -5c, calcium concentrations are uniformly low and roughly equal those in the Holocene. There are, however, striking differences between the appearance of MIS-5e at Summit and that at Vostok, Fig. 1. At Summit, there is a highly structured pattern of variation on both long (>1 kyr) and short (sub-century) timescales. By contrast, Vostok<sup>5</sup> shows smooth trends across the whole stage, which can only partially be explained by the coarser resolution of this record.

At Summit, only in the first part of MIS-5e are conditions mainly interglacial in character. Even here there is considerable fine structure with a series of transient returns to apparently cool interstadial conditions. This stage seems to correspond with the warmest stage of the Vostok profile where the temperature

\* M. Ankin, J. M. Barnola, J. Beer, T. Blunier, J. Chappellaz, H. B. Clausen, D. Dahl-Jensen, W. Dansgaard, M. De Angelis, R. J. Delmas, P. Duval, M. Fratta, A. Fuchs, K. Fuhrer, N. Gundestrup, C. Hammer, P. Iversen, S. Johnsen, J. Jouzel, J. Kipfstuhl, M. Legrand, C. Lorius, V. Maggi, H. Miller, J. C. Moore, H. Oeschger, G. Orombelli, D. A. Peel, G. Raisbeck, D. Raynaud, C. Schött-Hvidberg, J. Schwander, H. Shoji, R. Souchez, B. Stauffer, J. P. Steffensen, M. Stievenard, A. Sveinbjörnsdóttir, T. Thorsteinsson, E. W. Wolff.

FIG. 1 Comparison of profiles of stable isotopes and calcium ion concentrations through the last glacial cycle in central Greenland and Antarctica. a, 200 yr mean  $\delta^{18}\text{O}$ , from Summit<sup>7</sup>; b, 200 yr mean Ca, from Summit (based on 3-mm continuous measurements below 1,300 m depth, using continuous-flow analysis<sup>38</sup>); c,  $\delta\text{D}$  (1 m core-length average) from Vostok<sup>5</sup> with new timescale<sup>39</sup>; d, Ca ( $\sim 2$  kyr average) from Vostok<sup>16</sup>.



changes are believed to have preceded sea-level changes<sup>18</sup> by about 4 kyr. This sub-stage may represent less than 10 kyr of MIS-5e, more in accord with the  $\sim 11$ -kyr duration of MIS-5e derived from the SPECMAP curve<sup>12</sup>.

The climate over central Greenland seems to be highly sensitive to changes in circulation in the north Atlantic region. The GRIP<sup>7,9</sup> and GISP2<sup>10,19</sup> records, and previous Greenland deep-core records extending into the last glaciation<sup>8</sup>, all show evidence for series of 'mode switches' in the isotopic record in the late glacial period which are considered to reflect rapid shifts in the position of the Arctic frontal system, in turn believed to be connected with changes in ocean circulation<sup>20</sup>. Evidence for the most recent event<sup>21</sup>, the Younger Dryas (YD), has been traced in many independent media throughout areas under the influence of the north Atlantic, but it is a rather weak feature in Antarctic records<sup>22,23</sup>. There is also evidence for an abrupt transition at the end of the Eemian in pollen records from northeast France<sup>24,25</sup> and in coastal marine sediment sequences from western Norway<sup>26</sup> and off northwest Africa<sup>27</sup>. An isotopic record from an exposure of ancient ice at the margin of the Greenland ice sheet also appears to show structure in ice suggested to be of Eemian age<sup>28</sup>. It seems that the sharp termination of the Eemian and some of the structural features observed at the start of the Eemian may have parallels in other systems and in other regions.

### Climatic characteristics of the Eemian

Figure 2 shows the isotopic and dust (Ca) content, and electrical properties of the solid ice during the Eemian period. The low-frequency conductivity of the ice, measured by dielectric profiling (DEP<sup>29</sup>), is controlled by both the neutral salt (probably chloride) and the free acid concentration<sup>30</sup>. This parameter was closely paralleled by the d.c. electrical conductivity (ECM<sup>31</sup>) profile, reflecting only the total acidity along this section of the record. Chloride is mainly derived from sea spray, and concentrations observed in the ice are strongly regulated by atmospheric circulation. The acidic content of the ice is controlled by various sources (marine biogenic, sporadic volcanoes and  $\text{NO}_x$ ), and by transport and neutralization in the atmosphere. On the other hand, calcium levels strongly depend on a source that is known to have varied substantially throughout the course of the last glacial cycle—the extent of loess deposits<sup>32</sup> and of evaporites deposited on the continental shelves. Although we cannot at this stage separate out the various controls on these species, they undoubtedly have very different sensitivity to climate and circulation changes. We suggest that a similar pattern of composition should indicate broadly similar conditions of climate and atmospheric circulation.

In Fig. 2 we have identified three clear, warm substages of (MIS)-5e, which we have named 5e1, 5e3 and 5e5, apparent in

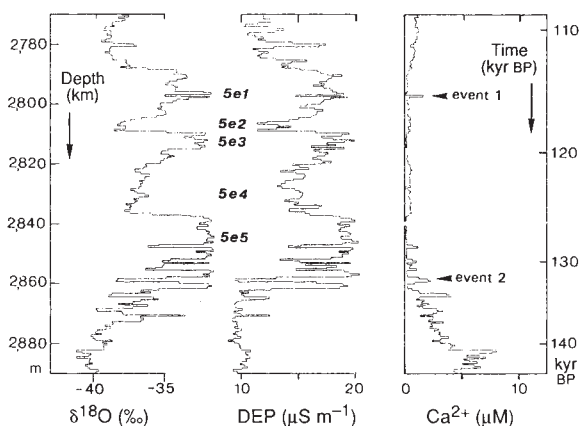


FIG. 2 Detailed profiles of selected isotopic and chemical constituents (averaged along 55 cm segments) through the Eemian (MIS-5e) section of the GRIP Summit core. MIS-5e is divided into three principal warm substages separated by up to 5 kyr of sustained cool periods, which we have named MIS-5e1 to 5e5. The profile for DEP<sup>40</sup> corresponds to the limiting high-frequency conductivity.



TABLE 1 Mean ice-core parameters for stages of the last climatic cycle

|                         | $\delta^{18}\text{O}$<br>(‰) | D excess<br>(‰) | Ca<br>( $\mu\text{M}$ ) | DEP<br>( $\mu\text{S m}^{-1}$ ) | ECM<br>(relative units) |
|-------------------------|------------------------------|-----------------|-------------------------|---------------------------------|-------------------------|
| Holocene                | -34.8                        | 8               | 0.203                   | 23.9                            | 1.43                    |
| Last Glacial<br>Maximum | -42                          | 6               | 10.11                   |                                 |                         |
| Stade                   | -41                          | 8               | 4.72                    | 11.01                           | 0.01                    |
| Interstade              | -38                          | 4               | 1.4                     | 12.6                            | 0.25                    |
| Warm Eemian             | -33.4                        | 8.5             | 0.28                    | 17.7                            | 1.07                    |
| Event 1                 | -39                          | 7               | 4.37                    | 11.36                           | 0.0054                  |
| Event 2                 | -38.4                        | 6               | 2.12                    | 9.66                            | 0.0036                  |
| 5E4                     | -36.8                        | 7               | 0.451                   | 14.75                           | 0.616                   |

Mean isotopic and chemical composition of ice deposited during the Eemian warm and cool stages in comparison with ice drawn from the principle climate stages of the last glacial cycle.

all constituents measured. In each of these, isotopically derived temperatures at least as high as in the Holocene levels were attained. Table 1 gives the comparative ice composition for different stages of the last glacial cycle. The isotopic and chemical signature of the warm stages of the Eemian is consistent with a similar atmospheric circulation to the Holocene. (Similar mean concentrations were obtained for Na, Mg, Cl,  $\text{SO}_4^{2-}$  and  $\text{NO}_3^-$ .) The difference in mean oxygen isotope ratios (1.4‰) suggests that temperatures were  $\sim 2^\circ\text{C}$  warmer than now during the warm stages of the Eemian (this is based on an isotope-temperature gradient for Greenland of  $1.5^\circ\text{C}$  per 1‰ change<sup>9</sup>). Deuterium excess ( $d = \delta\text{D} - 8 \times \delta^{18}\text{O}$ ) is sensitive to conditions in the moisture source region<sup>15</sup> and, in turn, to shifts in atmospheric circulation which could affect the stable isotope-temperature gradient. The similar values for mean deuterium excess in the Holocene and in the warm stages of the Eemian (Table 1) strengthen the idea that circulation changes are unlikely to have perturbed the gradients in the Eemian warm stages. Our dating of the record indicates that the unbroken warm stages characterize only 2 kyr of substage 5e5, <1 kyr of 5e3 and  $\sim 3$  kyr of 5e1 (except for one major transient event). The isotopic and chemical evidence suggests that conditions overall during the cool stage 5e4 were similar to the mid-glacial warm stages (interstadials),  $\sim 5^\circ\text{C}$  cooler than the Holocene. A marked antiphase variation between Ca concentration and  $\delta^{18}\text{O}$  during the last

glacial cycle has been reported previously in both Greenland and Antarctica<sup>17,33</sup>. This is reflected strongly in Table 1, where the relationship clearly extends through the sub-stages of the Eemian.

Although the isotope and chemical signature of ice laid down during the Eemian warm stages suggests that climate and circulation patterns were similar to those prevailing in the Holocene, at least in the north Atlantic region, the overall pattern of behaviour during MIS-5e is strikingly different. Figure 3 compares the frequency distribution of  $\delta^{18}\text{O}$  values for MIS-5 in comparison with those obtained for the Holocene and for the glacial sequence MIS-2 to 4. The marked bimodal behaviour of the mid-glacial sequence reflects the now well documented climate switching observed in all Greenland ice cores<sup>8</sup>. It seems that this behaviour was even more pronounced during MIS-5e, with a strongly defined intermediate mode. This contrasts with the single-mode isotope trend for the past 8 kyr of the Holocene, which seems to have a similar distribution to each of the component modes in the other stages, underlining the unusual stability of the present climate<sup>7</sup>.

The cool events between the principal substages of the Eemian indicate a marked change in climate in Greenland and adjacent regions, and given the sharp changes in chemical signature, these seem to be connected either with a sudden change in the larger-scale atmospheric circulation or with a sustained shift, for example in the position of the polar front, leading effectively to a systematic change in the average source regions of constituents in the atmosphere over central Greenland. Their persistence suggests that there may have been large-scale changes in oceanic circulation in the north Atlantic region. Measurements on the gas content of the ice should establish the global significance of these findings. Our first measurements of methane trapped in bubbles in the ice show concentrations of 620 p.p.b.v. and 650 p.p.b.v. at points within MIS-5e3 and 5e5 respectively, similar to the values reported for the last interglacial in the Vostok core<sup>34</sup>. A single point so far measured in the MIS-5e4 cool substage gives the significantly lower value of 530 p.p.b.v., at the lower limit of the noise envelope cited for the Vostok core for this period, and a value typical for an interstadial.

### Rapid climate change during the Eemian

The most striking feature of MIS-5e is a series of high-amplitude,

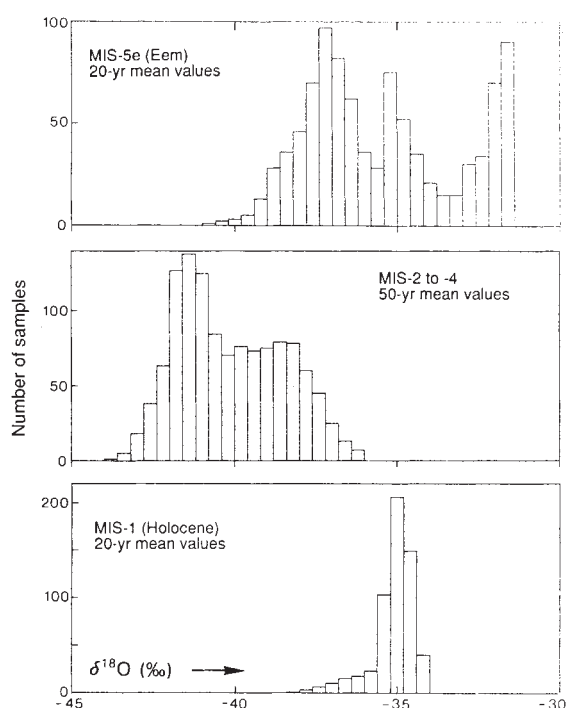


FIG. 3 Frequency distribution of the isotope data measured throughout MIS-5e in comparison with that observed during the last glaciation (MIS-2-4) and during the past 8 kyr of the Holocene. The distinctive single-mode characteristic of the Holocene contrasts with the bi- and trimodal behaviour observed through the last glaciation and during MIS-5e.

high-frequency oscillations seen in all constituents measured continuously along the core (Fig. 2). We have examined two contrasting events in greater detail. Event 1 (Fig. 4) is a catastrophic event within MIS-5e1 at the culmination of the Eemian, where the oxygen isotope values plunge to mid-glacial levels ( $-41\text{‰}$ ). The event is estimated to have lasted  $\sim 70$  years to judge from an annual layer thickness of 2.5 mm (estimated from the relationship between accumulation-rate<sup>7</sup> and  $\delta$ , and from flow modelling). Event 2 (Fig. 5) is one of the lengthy series of massive and sustained oscillations that marked the first  $\sim 8$  kyr of the Eemian and the end of the previous deglaciation sequence. Stable isotope values approach YD levels ( $\sim -40\text{‰}$ ) and were sustained for about 750 yr, comparable with the length of the YD itself. Maximum temperature decreases associated with these events are estimated at  $14^\circ\text{C}$  and  $10^\circ\text{C}$  respectively, on the basis of the suggested isotope-temperature relationship<sup>9</sup>. The deuterium excess is significantly reduced during both events, pointing to a lowering of associated moisture source temperatures. This suggests that any correction to the isotope-temperature gradient would only serve to increase the estimated temperature shifts.

The chemical signature of major species is broadly similar in both events. Strong negative correlation between  $\delta^{18}\text{O}$  and Ca evidently persisted in the more transient events within the Eemian, although peak Ca concentrations, similar to YD levels, are much lower than late glacial concentrations. Calcium concentrations in the warm stages either side of events 1 and 2 are roughly equal to Holocene levels.

Although many features of the Eemian cool events seem to parallel the changes in the YD, there is a striking difference in the signal for deuterium excess which is strongly in phase with  $\delta^{18}\text{O}$  throughout the Eemian, including events 1 and 2. During the late glacial and including the YD event, these parameters are strongly in antiphase<sup>21</sup>. It seems clear that although a similar mode switching in ocean-atmosphere circulations may be implicated in both the Eemian and late glacial events (including YD), there has been a fundamental difference in the oceanic source regions involved. The vast production of melt water from decaying ice sheets could have been involved in the YD, but was not available at least for event 1, which occurred in the final part of the Eemian.

There is some indication in event 1 of a catastrophic start to the event with a more tapered recovery. This can be seen in the dust profile, a parameter that suffers least from smoothing by diffusion and has been measured at millimetre resolution. Surging of (for example) the west Antarctic ice sheet could have provided the necessary trigger to the system<sup>35</sup>, but the rapidity of the start of the event as seen in Greenland would be surprising, especially as it seems to have passed unrecognized in the Vostok

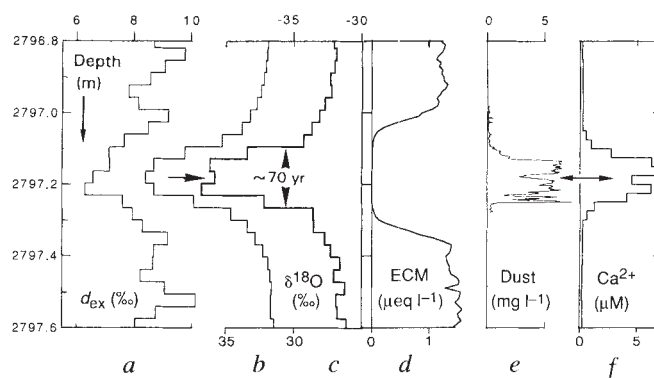


FIG. 4 Profiles of five parameters through 'event 1', a rapid climatic oscillation ( $\sim 70$  yr duration) at the culmination of the Eemian interglacial,  $\sim 115$  kyr BP. a, Deuterium excess<sup>15</sup>; b, oxygen isotope ratio<sup>7</sup>; c, same as b, but deconvoluted to account for diffusion (estimated diffusion length 3 cm); d, acidity measured by ECM in microequivalents per litre<sup>31</sup>; e, dust concentration measured from scattered laser light and calibrated by Coulter Counter by integrating size distribution<sup>33</sup>; f, calcium ion concentration<sup>38</sup>.

core. Alternatively, a sudden weakening of the north Atlantic current could have allowed the Greenland current to carry icebergs much further south. A general cooling of the north Atlantic associated with such events could, in part, account for the observed lower deuterium excess. Whatever the 'trigger' for these events, their duration of up to several hundred years is most probably connected with a re-ordering of the oceanic circulation. Such features have not, however, been observed through the Holocene, despite important climate changes during this period—these include the Climatic Optimum and the Little Ice Age.

We have considered the possibility that the events may have been generated or sharpened by processes occurring within the ice sheets, but several factors argue strongly against this. First, the events are located  $\sim 200$  m above bedrock. Visible examination of the core revealed no evidence for significant disturbance of the layering down to 2,847 m, or  $\sim 50$  m below the depth of event 1, although examination of cloudy bands showed that the layering was becoming inclined by up to  $21^\circ$  by the depth of event 2. Although this might indicate some migration of the position of the ice divide since the Eemian<sup>7</sup>, no indication of any discontinuity in the record was observed after the start of the

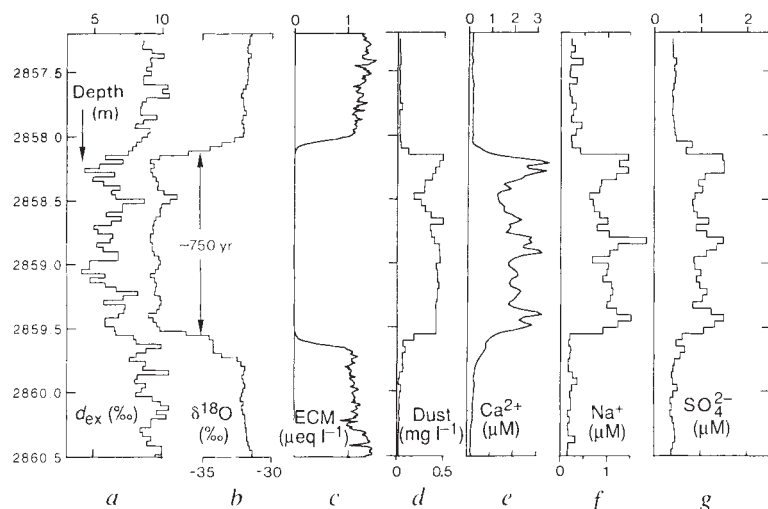


FIG. 5 Profiles of oxygen isotope ratio<sup>7</sup>, calcium ion concentration<sup>38</sup>, acidity by ECM<sup>31</sup>, dust<sup>33</sup>, deuterium excess<sup>15</sup>, sodium and sulphate<sup>41</sup> through 'event 2', a climate oscillation sustained for  $\sim 750$  yr in the initial stages of the Eemian interglacial  $\sim 131$  kyr BP.



Eemian. Observations on the core have revealed a strong anticorrelation between crystal size and dust or Ca content, with sharp changes at the boundaries of both events 1 and 2. Studies on the ice adjacent to and within event 1 show a significant strengthening of the fabric (pattern of preferred orientation) within the smaller crystals inside the event. It is not clear how far this could influence the rate at which the ice thins. Patterson<sup>36</sup> has argued that the effect should not be important near an ice divide where little shearing is expected.

Second, transition times at the start and at the end of the events are similar to those observed for the mid- and late-glacial events, even allowing for peak broadening due to diffusion. Some diffusion is indicated by the symmetry and broader appearance of the ECM ( $H^+$ ), DEP and  $\delta^{18}O$  profiles with respect to the dust profiles. If dust is considered to be immobile, and it is also assumed the shifts were originally simultaneous, then we can estimate a mean diffusion length of  $\sim 8$  cm for both  $H^+$  and  $\delta^{18}O$ . This can be compared with a diffusion length for ice molecules of 3–4 cm calculated on the basis of a mean ice temperature of  $-15^\circ C$ , a vertical strain rate of  $3 \times 10^{-5} \text{ yr}^{-1}$  estimated for the Eemian ice, and a self-diffusion coefficient for ice of  $2.5 \times 10^{-8} \text{ m}^2 \text{ yr}^{-1}$ . If we therefore assume that the dust signal represents the broad shift in atmospheric circulation pattern at the boundaries of these events, this suggests that the shift was largely completed within  $\sim 10$  years for event 1 and  $\sim 30$  years for event 2. These rates of change are comparable with the rates indicated previously for both YD termination<sup>21</sup> and for the late glacial events<sup>7,8</sup>.

At this stage, estimates of the rate of change of climate parameters must be tentative, but such processes cannot affect the amplitude of the events themselves. The overall picture of an atmospheric circulation system oscillating between two well defined modes is unaffected.

## What the ice cores tell us

Although we knew that abrupt and large climatic oscillations could occur during the cold stages of the glacial-interglacial cycle, the oscillations seen during a warm stage are new. They are prevalent in the early to mid-stages of the warmest part of the Eemian interglacial but occur throughout MIS-5e. The mode switches may be completed in as little as 1–2 decades and can become latched for anything between 70 yr and 5 kyr. The signa-

ture of these events revealed in the chemical profiles in the GRIP core is consistent with large changes in atmospheric circulation patterns in the north Atlantic, and important climate changes at least over Greenland and adjacent regions are indicated. The length of these events indicates that shifts in ocean circulation were involved, as is believed to have been the case with the YD event, so the effects are likely to have been much more widespread.

Given the history of the last 150 kyr, the past 8 kyr has been strangely stable; only during the final  $\sim 2$  kyr of the warmest stage of the Eemian interglacial do our data demonstrate a similar period of stability. The unexpected finding that the remainder of the Eemian period was interrupted by a series of oscillations, apparently reflecting reversals to a 'mid-glacial' climate is extremely difficult to explain. Perhaps the most pressing question is why similar oscillations do not persist today, as the Eemian period is often considered as an analogue for a world slightly warmer than today's. It has been suggested that the north Atlantic Oscillation or related processes may have a bearing on decadal-frequency features<sup>10,37</sup>, as they are linked to strong interannual variability at coastal Greenland stations, but clearly this has not left significant evidence in the Holocene ice-core records. If such an oscillation caused the earlier features, it must have been far stronger. Moreover we need to explain how a mode switch can be sustained for 70 yr—5 kyr after being induced by, perhaps, a higher frequency switching in the atmospheric system, or by some catastrophic event such as ice-sheet surging.

As the GISP2 drilling nears bedrock, we can rapidly expect both confirmation of these findings and new clues as the sister cores are subjected to closer scrutiny. The analysis of new parameters should allow us to narrow down the range of possible changes in ocean and atmosphere that favoured instability in the climate. Coordinated analysis of the two cores should allow much firmer dating than has hitherto been possible in any single ice-core, and remove beyond reasonable doubt any question relating to local interferences on the records. Jointly, they may provide the vital clue to answer the question of what it might take to induce Eemian-type instability into the Holocene pattern of climate. Man is already perturbing one of the factors that may be involved, the greenhouse gases, and our first tentative measurements indicate that they may be linked to climatic changes within the Eemian.  $\square$

Received 5 April; accepted 3 June 1993.

1. IGBP Global Change, 12, IGBP: A Study of Global Change. The Initial Core Projects (International Geosphere-Biosphere Programme, 1990).
2. Dansgaard, W. in *The Climate of Europe: Past, Present and Future* (eds Flohn, H. & Fantechi, R.) 208–225 (Reidel, Dordrecht, 1984).
3. Anderson, P. et al. *Quat. Int.* **10–12**, 9–28 (1991).
4. Müller, H. in *Man's Impact on Climate* (eds Bach, W., Pankreth, J. & Kellogg, W.) 29–41 (Elsevier, Amsterdam, 1979).
5. Jouzel, J. et al. *Nature* **329**, 403–408 (1987).
6. Dansgaard, W. et al. *Science* **218**, 1273–1277 (1982).
7. Dansgaard, W. et al. *Nature* **364**, 218–220 (1993).
8. Oeschger, H. & Arquit, A. in *Global Changes of the Past* (ed. Bradley, R. S.) 175–200. (UCAR/Office for Interdisciplinary Earth Studies, Boulder, Colorado, 1991).
9. Johnsen, S. J. et al. *Nature* **359**, 311–313 (1992).
10. Taylor, K. C. et al. *Nature* **361**, 432–436 (1993).
11. Letréguilly, A., Reeh, N. & Huybrechts, P. *Global Planet. Change* **90**, 385–394 (1991).
12. Martinson, D. G. et al. *Quat. Res.* **27**, 1–29 (1987).
13. Winograd, I. J. et al. *Science* **258**, 255–260 (1992).
14. Dansgaard, W. *Tellus* **18**, 436–468 (1964).
15. Johnsen, S. J., Dansgaard, W. & White, J. W. C. *Tellus* **B41**, 452–468 (1989).
16. Legrand, M. R., Lorius, C., Barkov, N. I. & Petrov, V. N. *Atmos. Environ.* **22**, 317–331 (1988).
17. Delmas, R. J. & Legrand, M., in *Dahlem Konferenzen, The Environmental Record in Glaciers and Ice Sheets* (eds Oeschger, H. & Langway, C. C. Jr) 319–341 (Wiley, Chichester, 1989).
18. Sowers, T., Bender, M., Raynaud, D., Korotkevich, Y. S. & Orchardo, J. *Paleoceanography* **6**, 679–696 (1991).
19. Alley, R. B. et al. *Nature* **362**, 527–529 (1993).
20. Broecker, W. S., Bond, G., Klas, M., Bonani, G. & Wolff, W. et al. *Paleoceanography* **5**, 469–477 (1990).
21. Dansgaard, W., White, J. W. C. & Johnsen, S. J. *Nature* **339**, 532–534 (1989).
22. Jouzel, J., Lorius, C., Meriviat, L. & Petit, J.-R. in *NATO ASI Series C*, 216, *Abrupt Climatic Change* (eds Berger, W. H. & Labeyrie, L. D.) 235–245 (Reidel, Dordrecht, 1987).
23. Jouzel, J. et al., in *NATO ASI Series 12, The Last Deglaciation: Absolute and Radiocarbon Chronologies* (eds Bard, E. & Broecker, W. S.) 229–226 (1992).
24. Woillard, G. *Nature* **281**, 558–562 (1979).
25. Guiot, J., Pons, A., de Beaulieu, J. L. & Reille, M. *Nature* **338**, 309–313 (1989).
26. Mangerud, J., Sørensen, E. & Sejrup, H.-P. *Nature* **227**, 189–192 (1979).
27. Eglinton, G. et al. *Nature* **356**, 423–426 (1992).
28. Reeh, N., Oerter, H., Letréguilly, A., Miller, H. & Hubberten, H.-W. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **90**, 373–383 (1991).
29. Moore, J. C., Mulvaney, R. & Paren, J. G. *Geophys. Res. Lett.* **16**, 1177–1180 (1989).
30. Moore, J. C., Paren, J. G. & Oerter, H. J. *Geophys. Res.* **97**, 19803–19812 (1992).
31. Hammer, C. U. J. *phys. Chem.* **87**, 4099–4103 (1983).
32. Pye, K. *Aeolian Dust and Dust Deposits* (Academic, London, 1987).
33. Hammer, C. U. et al. in *Greenland Ice core: Geophysics, Geochemistry and the Environment* (eds Langway, C. C. Jr, Oeschger, H. & Dansgaard, W.) 90–94 (Am. geophys. Un. Geophys. Monogr. 33, Washington DC, 1985).
34. Chappellaz, J., Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **345**, 127–131 (1990).
35. Hollin, J. T. *Boreas* **6**, 33–52 (1977).
36. Paterson, W. S. B. *Cold Reg. Sci. Technol.* **20**, 75–98 (1991).
37. Lehman, S. *Nature* **361**, 404–405 (1993).
38. Fuhrer, K., Neftel, A., Anklin, M. & Maggi, V. *Atmos. Environ.* (in the press).
39. Jouzel, J. et al. *Nature* (in the press).
40. Moore, J. C., Wolff, E. W., Clausen, H. B. & Hammer, C. U. J. *Geophys. Res.* **97**, 1887–1896 (1992).
41. Steffensen, J. P. *Ann. Glaciol.* **10**, 171–177 (1988).

ACKNOWLEDGEMENTS. This work is a contribution to the Greenland Ice-core Project (GRIP), co-ordinated and supported by the European Science Foundation. We thank the national funding agencies and organizations in Belgium, Denmark, France, Germany, Iceland, Italy, Switzerland and the United Kingdom together with the XII Directorate of CEC for financial support. We thank the many individuals who have supported this project.

# Molecular basis of the *little* mouse phenotype and implications for cell type-specific growth

Sheng-Cai Lin\*, Chijen R. Lin\*, Ilya Gukovsky\*, Aldons J. Lusis†, Paul E. Sawchenko‡ & Michael G. Rosenfeld\*

\* Howard Hughes Medical Institute, Eukaryotic Regulatory Biology Program, School and Department of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0648, USA

† Department of Microbiology and Molecular Genetics, Department of Medicine and Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California 90024, USA

‡ Laboratory of Neuronal Structure and Function, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

**The molecular basis for the *little* (*lit*) mouse phenotype, characterized by a hypoplastic anterior pituitary gland, is the mutation of a single nucleotide that alters Asp 60 to Gly in the growth hormone releasing factor receptor. Detailed analysis of the *lit* mouse anterior pituitary reveals spatially distinct proliferative zones of growth hormone-producing stem cells and mature somatotrophs, each regulated by a different trophic factor. This sequential growth factor requirement for a specific cell type may exemplify a common strategy for regulating cellular proliferation in other mammalian organs.**

THE mammalian anterior pituitary gland is a well developed model for investigating the mechanisms by which cells arising from a common primordium proliferate and terminally differentiate to generate a specialized organ containing functionally diverse cell phenotypes<sup>1,2</sup>. In addition to somatotrophs, lactotrophs, thyrotrophs, gonadotrophs and corticotrophs, a small percentage of anterior pituitary cells, referred to as somatomammotrophs, coexpress growth hormone and prolactin<sup>3</sup>. Initial proliferation occurs preferentially in the ectodermal cells in direct contact with the neuroectoderm, with specific cell phenotypes emerging in a distinct temporal and spatial fashion during ontogeny<sup>4</sup>, subsequently becoming intermingled.

The tissue-specific POU domain protein, Pit-1 (refs 5, 6), initially expressed in the developing murine anterior pituitary gland on e14 (ref. 4), serves as a critical activator of the prolactin and growth hormone genes<sup>5-8</sup>. It is ultimately expressed in mature lactotrophs, somatotrophs and thyrotrophs<sup>4,5,9</sup>, the three cell types absent in the hypoplastic pituitary gland of the Pit-1-defective Snell dwarf mouse (*dw*)<sup>10-15</sup>. A second genetically transmitted dwarfism in mice, referred to as *little* (*lit*/*lit*)<sup>15-19</sup>, is characterized by anterior pituitary hypoplasia, a marked decrease in glandular growth hormone messengerRNA and protein, and some decrease of prolactin.

Somatotroph cell proliferation, as well as growth hormone gene expression and secretion, is stimulated by increases in intracellular cyclicAMP levels, regulated by the hypothalamic growth hormone releasing factor (GRF) which is delivered by a portal circulatory system<sup>20-32</sup>. The pituitary effects of a constitutively active form of G<sub>s</sub>  $\alpha$ -subunit<sup>20</sup>, or targeted expression of cholera toxin<sup>25</sup>, GRF<sup>29</sup>, or dominant negative mutants of the cAMP-response element binding protein (CREB)<sup>30</sup> are in accord with a putative proliferative role of cAMP in somatotrophs. Thus, somatotrophs are one of the relatively limited number of cell types in which proliferation is positively regulated by increased intracellular levels of cAMP<sup>33-38</sup>. The observations that growth hormone secretion can be stimulated in *lit* mice by pharmacological agents that increase intracellular levels of cAMP<sup>17</sup>, and the failure to activate GRF receptor gene expression in the Pit-1-defective Snell dwarf<sup>12</sup>, suggest that a defect in the GRF receptor could account for the pituitary dysfunction in the *lit*/*lit* phenotype.

Here we define the molecular basis of the *lit*/*lit* dwarf mouse as a mutation in the GRF receptor gene. Sequential, spatially restricted requirements for distinct trophic factors underlie the continuum of proliferating growth hormone-producing stem cells and committed somatotrophs, with GRF-dependent stimulation of intracellular cAMP levels required to generate the caudomedial complement of proliferating somatotrophs.

## GRF receptor gene mutation in the *lit* mouse

Based on identification of the GRF receptor complementary DNA<sup>12,39,40</sup>, its encoding gene could be mapped in the mouse by linkage analyses using an interspecific backcross between the laboratory strain C57BL/6J and the interfertile species *Mus spretus*, with C57BL/6J as the recurrent parent. A restriction fragment length variant (RFLV) of the gene was identified by Southern hybridization using a full-length GRF receptor cDNA probe. After cleavage of genomic DNA with *Hind*III, strain C57BL/6J had a single 13.0 kilobase (kb) band. Strain *M. spretus* had 10.3 and 5.6 kb bands and F<sub>1</sub> mice had all three bands (data not shown). Based on the segregation pattern of the RFLV scored in 66 backcross mice previously typed for over 200 markers spanning the mouse genome, linkage was apparent for several markers located in the proximal region of mouse chromosome 6, D6Nds4 and D6Mit8 being the nearest flanking markers (Fig. 1a). Within the error of determination of linkage distances, the location of the GRF receptor gene coincided with that of the *lit*/*lit* mutation<sup>15,49</sup>; however, no obvious rearrangements or deletions in the GRF receptor genomic locus could be detected by Southern blot analyses (data not shown).

The GRF receptor gene was analysed from multiple genomic clones that encompassed the GRF receptor coding sequence from *lit*/*lit* and wild-type genomic DNA libraries. The GRF receptor gene encompassed >10 kb, with 13 exons represented in the coding region, as well as an alternatively used exon<sup>12,39</sup> located between exons X and XI (Fig. 1c, and data not shown). All exons, introns and >2.5 kb 5' and 3' flanking information were sequenced, revealing a single significant nucleotide alteration (A→G) in the second nucleotide of codon 60, predicting an encoded glycine residue instead of an aspartic acid (Asp 60→Gly) (Fig. 1b, c). As shown in Fig 1b, the aspartic acid residue is conserved in all known members of the GRF receptor



class of seven-transmembrane helix,  $G_s$  protein-coupled receptors<sup>41–46</sup> and a nonconservative mutation to glycine in this position might be predicted to alter protein structure significantly in the N-terminal domain of the receptor. This mutation was confirmed from the sequence of polymerase chain reaction (PCR) amplified wild-type and *lit/lit* genomic DNA, and by sequencing four wild-type and three *lit/lit* independent genomic clones. Therefore, the (Asp 60→Gly) mutation provided a potential explanation for the *lit* mouse phenotype.

To test this, pools of permanently transfected cells were generated that expressed the wild-type or *lit* GRF receptor, and the ability of GRF to transfer signals effectively through the wild-type and *lit* GRF receptor was addressed by assessing intracellular cAMP levels. Although addition of GRF effectively increased levels of cAMP in heterologous cells expressing the wild-type GRF receptor<sup>12,39,40</sup>, stimulation of intracellular cAMP could not be detected in cells expressing the Asp 60→Gly *lit* GRF receptor, even at GRF concentrations as high as  $10^{-6}$  M (Fig. 2). These data suggested that the GRF receptor encoded in the *lit/lit* mouse was functionally defective and unable to transduce GRF-dependent increases in intracellular cAMP levels. This failure of *lit* GRF receptors to mediate signal transduction was confirmed by the *lit* Asp 60→Gly receptor mutant being unable to mediate GRF-dependent transcriptional stimulation of promoters containing cAMP response elements (data not shown). Therefore, the molecular basis of the *lit* mouse phenotype

appears to be a point mutation in the N-terminal ligand-binding domain of the GRF receptor that results in a failure of GRF-dependent, receptor-mediated signal transduction.

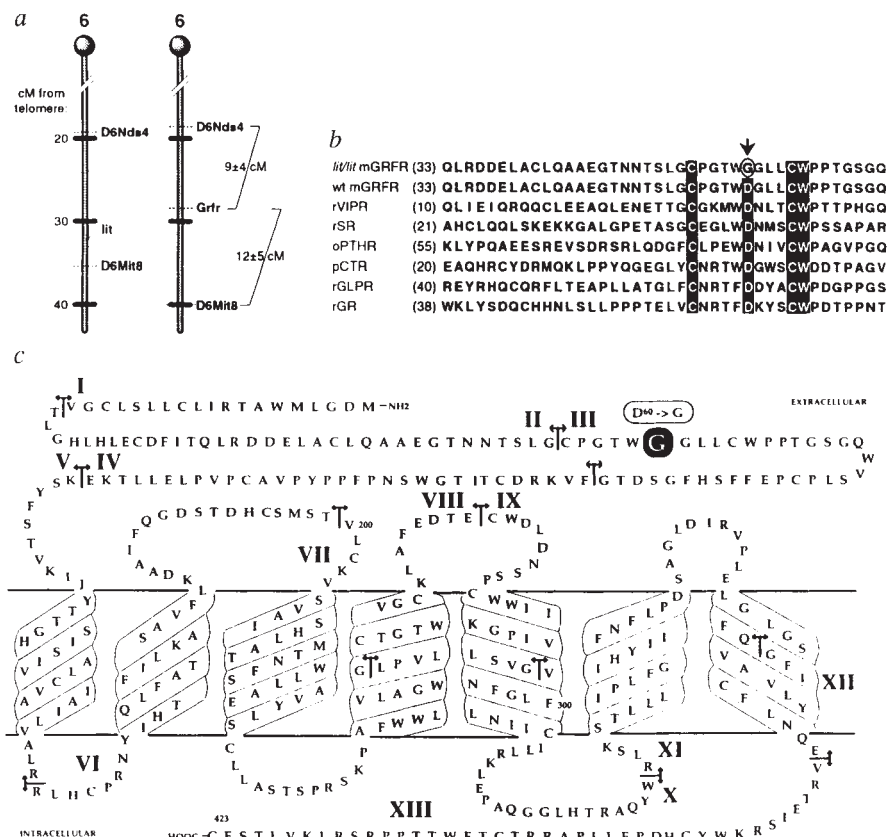
## Anterior pituitary cellular proliferation events

The *lit* mouse provided a favourable model for investigating the role of a specific receptor in the regulation of cellular proliferative events in mammalian organogenesis. Initially, ontogeny of anterior pituitary gland development in *lit* mice was analysed using *in situ* hybridization. Two eighteen-base-pair oligonucleotides encompassing the point mutation were used for genotyping<sup>12</sup>. Analysis of the initial activation of Pit-1, growth hormone, GRF receptor gene expression and the size of the developing anterior pituitary, on e16, e17 and e18 (e16 and e18 are not shown) were indistinguishable between wild-type and *lit/lit* mice, indicating the preservation of normal proliferative events during ontogeny (Fig. 3).

But a striking hypoplasia was clearly apparent in the mature gland. *In situ* hybridization analysis revealed that this hypoplasia reflected a preferential loss of growth hormone-expressing somatotrophs when compared to normal pituitaries (Figs 4 and 5). A more detailed analysis revealed an unexpected spatial distribution of the remaining somatotroph cells. Aggregates of growth hormone-expressing cells were predominantly located toward the periphery of the mature gland, with clusters of cells effectively expressing high levels of growth hormone consistently

FIG. 1 Identification of the *lit/lit* defect as an A→G nucleotide transition in the GRF gene. **a**, Consensus map of mouse chromosome 6, containing markers D6Nds4 and D6Mit8 as well as the *lit* gene, is shown at the left, with the rough distance (cM) from the centromere indicated<sup>49</sup>. The linkage map generated from analysis of 67 backcross mice, including the GRFR gene, is shown at the right. Distances between markers are expressed as cM ± the standard error<sup>52</sup>. **b**,  $\lambda$ FixII library from *lit/lit* genomic DNA was screened and seven genomic clones were isolated, one of which encompassed the entire gene and was completely sequenced. Sequencing all *lit/lit* exons revealed a single alteration in coding information, an A→G transition in the second position of codon 60, altering the encoded amino acid aspartic acid to glycine, as shown, involving an amino acid conserved in the N-terminus of all the members of this class of seven-transmembrane helix,  $G_s$  protein-coupled receptors. Receptors are: vasoactive intestinal peptide (VIPR), secretin (SR), parathyroid hormone (PTHr), calcitonin (CTR), glucagon-like peptide (GLPR), and glucagon (GR); r, rat, o, ovine; p, porcine. **c**, Thirteen exons were used to generate a mature mouse GRF mRNA, as shown; at least one additional, potential exon is located between exons X and XI, which has been established to be alternatively spliced in rat and human<sup>12,39</sup>.

**METHODS.** The construction of the (C57BL/6J × *Mus spretus*)  $F_1$  × C57BL/6J interspecific backcross mice was as previously described<sup>53</sup>. The backcross animals were typed for over 200 microsatellite and RFLV genetic markers, including all chromosomes. Southern analysis was essentially as previously described at 65 °C using a full-length GRF receptor cDNA probe labelled by random priming to specific activity of about  $10^9$  c.p.m.  $\mu\text{g}^{-1}$ . Ten different enzymes were examined for RFLVs and *Hind*III was chosen because it produced clearly distinct hybridization patterns. Microsatellite polymorphisms were typed using PCR amplification followed by separation of the radiolabelled products on acrylamide gels as described<sup>54</sup>. Linkage maps were constructed and statistically analysed



using the MAPMAKER program<sup>55</sup>. A GRFR cDNA fragment (1.8 kb) was used as a probe to screen wild-type and *lit/lit* mouse  $\lambda$ FixII DNA libraries ( $\sim 1.5 \times 10^6$  plaques from each were screened). Eleven positive clones and seven positive clones, respectively, were purified, subcloned, and inserts sequenced using the dideoxy nucleotide chain termination method<sup>56</sup>. The mutation was confirmed in three independent *lit/lit* clones that encompassed the putative mutation. Four wild-type clones were analysed, none of which contained the *lit* mouse GRFR alteration.

localized to the anterolateral aspects of the gland (Figs 4 and 5). Very few cells giving a signal with the growth hormone probe were observed in the caudomedial portion of the anterior pituitary gland (Fig. 4). This pattern was independently confirmed by a histochemical analysis using anti-growth hormone sera, revealing aggregates of growth hormone-containing cells largely confined to the lateral aspect of the gland (Fig. 5). The pattern of GRF receptor gene expression essentially corresponded to that observed for growth hormone, with some cells exhibiting high levels of expression in the anterolateral portion of the gland, but without detectable expression in the caudomedial aspect, consistent with the inability of somatotrophs to proliferate into this area (Fig. 4).

In contrast, using an antisense prolactin probe, lactotrophs were uniformly distributed over the entire anterior pituitary gland (Fig. 4) at apparently normal density, indicating that the *lit/lit* GRF receptor defect did not markedly impair the progression of the lactotroph phenotype. In light of the decreased gland size, the number of lactotrophs was actually likely to be slightly decreased. Together, these data suggested that the putative common precursor of lactotroph and somatotroph cell types<sup>47,48</sup> was unaffected by the *lit* GRF receptor defect. Agreeing with these findings, *pit-1* signals were similarly distributed over the entire gland, including the caudomedial aspect (Fig. 5), consistent with its known expression in lactotrophs and thyrotrophs<sup>4</sup>. The detectable expression of the POMC gene transcript, a marker of corticotrophs and the intermediate lobe, was similar in normal and *lit/lit* anterior pituitaries (Fig. 4), but appeared to be more

dense in the *lit/lit* gland due to the decreased number of somatotrophs. We could find no evidence of an increase in the percentage of the somatomammotroph cells in the *lit* mouse by concurrent dual indirect immunofluorescence using anti-prolactin and anti-growth hormone sera (Fig. 5 and data not shown), suggesting that a normal progression to somatotroph and lactotroph stem cells was maintained.

### Pit-1 regulation of the GRF receptor promoter

Analysis of the Pit-1-defective Snell dwarf mouse that has a marked pituitary hypoplasia, revealed a failure to activate GRF receptor or growth hormone gene expression<sup>12</sup>. Cotransfection of fusion genes under control of the GRF receptor promoter with Pit-1 expression vectors produced a consistent stimulation of GRF receptor reporter gene expression in heterologous cell types, suggesting that the GRF receptor is likely to be a direct Pit-1 target gene (Fig. 6a). The defective *dw* Pit-1 protein was unable to transactivate GRF receptor fusion genes (Fig. 6a). Pit-1-dependent activation of GRF receptor promoter was only minimally altered (~25%) by the human Ala 158→Pro Pit-1 mutant protein (data not shown), in contrast to the virtual elimination of Pit-1-dependent activation of the prolactin enhancer/promoter<sup>50</sup>. Although the precise basis for promoter-specific differences in response to the Ala 158→Pro Pit-1 mutant<sup>50</sup> remains to be defined, these data suggest that the molecular basis for the maintenance of pituitary gland size in the individuals harbouring the Ala 158→Pro mutation may, in part, reflect fundamental differences between the nature of Pit-1 response elements in growth hormone and prolactin genes and in the GRF receptor.

### Discussion

We have exploited a genetic defect in anterior pituitary development as a tool for studying the molecular mechanisms that underlie cell type-specific proliferation, as well as terminal differentiation, of distinct cell phenotypes. In view of previously reported physiological, pharmacological and transgenic data linking somatotroph proliferation to increased levels of cAMP<sup>20-30</sup>, the demonstration that the *lit* mouse encodes a defective GRF receptor provides direct evidence that a single cell-specific receptor is physiologically required for the proliferation events necessary to generate a full complement of somatotrophs. A defective GRF receptor results in about a 10-fold decrease in somatotroph cells and in growth hormone production<sup>15, 19</sup> compared to those in normal pituitary glands. A detailed analysis of pituitary gland development in the *lit* mouse has also revealed that, despite the apparent intermixing of all cell types in the mature anterior pituitary gland, there is retained a spatial polarity of cellular proliferation and a temporal order in which specific growth factors are required to maintain proliferation events. The *lit* GRF receptor defect exerts no apparent effects during early anterior pituitary development, because a full initial complement of Pit-1, growth hormone and GRF receptor-positive cells is present. Only after pituitary ontogeny does a defect in somatotroph proliferation manifest in the *lit/lit* mouse. But even in the mature gland the GRF receptor is still not required for proliferation of the stem cells that ultimately give rise to mature somatotrophs and lactotrophs.

We suggest, as in the model in Fig. 6b, that a stable population of growth hormone-expressing cells localized to the anterolateral periphery of the *lit/lit* gland, represents a GRF-independent 'mitotic zone', in which presomatotroph-stem and somatotroph-stem cell populations arise and are maintained, based on their response to other modulators of proliferation (growth factors). Analogous to the 'mitotic zone' used during ontogeny of the anterior pituitary<sup>31</sup>, a zone of stem cell proliferation is, therefore, apparently maintained in the mature gland, although mitosis of somatotrophs continues across the entire gland<sup>51</sup>. Therefore, somatotrophs can be proposed to divide sequentially from the anterolateral region toward the caudomedial aspects of the gland

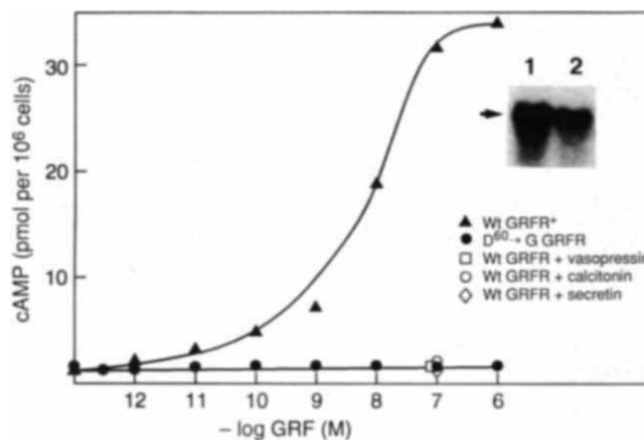


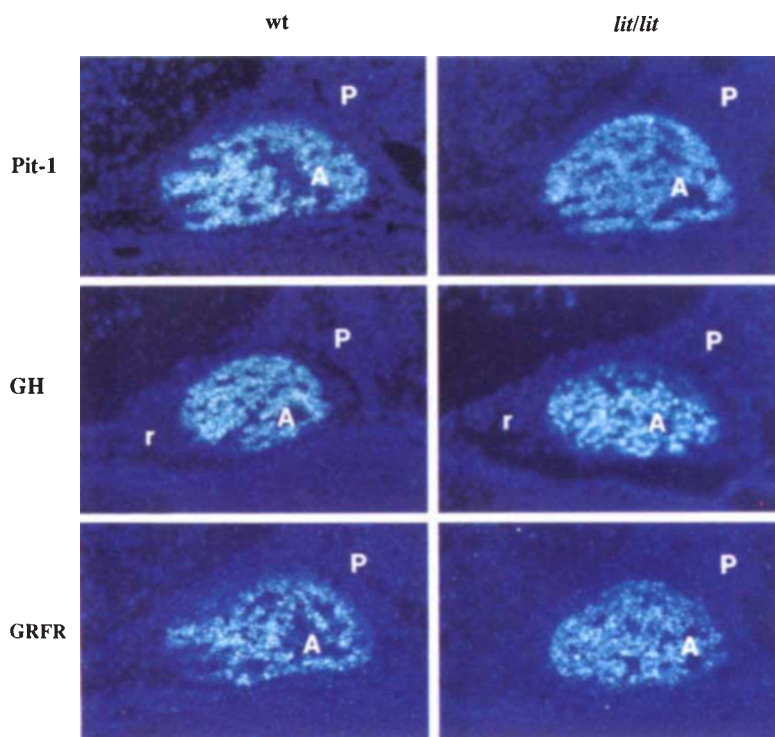
FIG. 2 The mutant *lit/lit* GRF receptor is defective in transducing GRF-dependent increase in intracellular cAMP levels. Permanent non-clonal CV1 cell lines were transfected with a pCEP4-GRFR expression vector, or the vector alone, and effects of hormones on intracellular cAMP levels were determined. Results are the average of triplicate determinations differing by <5%. Stimulation was GRF-dependent, and was not observed in response to GRF in control transfected cells, or to  $10^{-7}$  M corticotrophin releasing factor (CRF), secretin or calcitonin in GRF receptor-expressing cells. Insert shows northern blot from *lit* (lane 1) or wild-type (lane 2) transfectants probed with GRF receptor cDNA. The arrow indicates the reactive RNA band migrating at the expected size. Similar results were obtained in three additional experiments of similar design.

**METHODS.** The mouse GRF receptor D60→G mutant receptor was generated by site-directed mutagenesis (using ACAGCAGCCACCCCAGGTCC as the mutagenic oligonucleotide)<sup>5</sup>. GRF receptor cDNA was placed under control of the CMV promoter in the pCEP4 vector (Invitrogen) and used to generate permanently transfected cell lines, with hygromycin as the selection agent. Exponentially growing CV1 cells (100 mm dishes) were transfected with pCEP4-GRFR expression plasmid DNA (10 µg) using a calcium phosphate method<sup>5</sup>. Stable cells were plated to confluences of 30–50% and cAMP assays were as previously described<sup>12</sup>. Values are normalized.



FIG. 3 Ontogeny of anterior pituitary development in the *lit/lit* mouse. Anterior pituitary glands of e16, e17 and e18  $+ / lit$  mice (referred to as wild-type (WT)) and littermates (referred to as *lit/lit* mice) were subjected to *in situ* hybridization histochemistry using antisense growth hormone, GRF receptor (GRFR) and Pit-1 probes, as previously described<sup>12</sup>. Results on e17.5 embryos are shown. Signals were all confined to the central-caudal portion of the anterior pituitary (A). P, posterior lobe; r, rostral tip. No specific difference between wild-type and *lit/lit* signals were detected on e16, e17, e18 glands for any probe described above.

METHODS. Embryos of  $+ / lit$  and *lit/lit* pregnant females plugged by  $+ / lit$  males were fixed in a 4% paraformaldehyde solution and serially sectioned at 10  $\mu$ m. Hybridizations were done using  $5 \times 10^6$  c.p.m.  $ml^{-1}$   $^{35}S$ -labelled antisense (or sense) RNA probes of GH, Pit-1 or GRFR, as previously described<sup>4,12</sup>. After hybridization sections were washed, dehydrated, and exposed to X-ray film (Amersham). After autoradiography, slides were defatted, stained using bis-benzamide, coated with emulsion (NTB-2), and developed after 2–14 days of exposure. No specific hybridization was detected using comparable levels of sense-strand RNA probe. *lit/lit* homozygotes were identified by dot blot and Southern hybridization of PCR-amplified genomic DNA with probes spanning the mutated nucleotide (wt: TGGGACCTGGGA-TGGGCT; *lit/lit*: TGGGACCTGGGGTGGGCT).

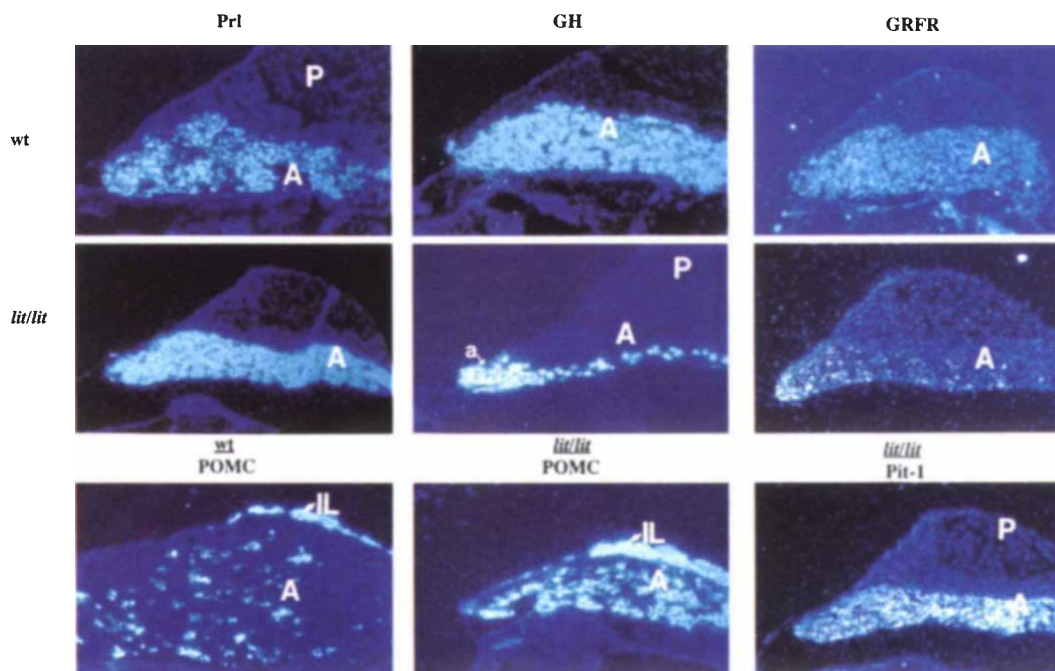


in a fashion that is completely dependent on GRF, based on the effects of the absence of functional GRF receptors in the *lit/lit* homozygote. These data are consistent with the previous observations in transgenic mice expressing a dominant-negative CREB protein under control of the growth hormone promoter, in which the pituitary contains residual somatotrophs<sup>30</sup>, whereas

all somatotrophs were depleted in the gangcyclovir-treated mature GH-TK or the GH-diphtheria toxin transgenic mouse pituitary<sup>48,49</sup>. Together, these data suggest that initial somatotroph stem cell proliferation is not likely to be under control of a cAMP-dependent signal transduction system. As somatotrophs continue to proliferate centrally, GRF becomes required for con-

FIG. 4 The *lit* mouse has a

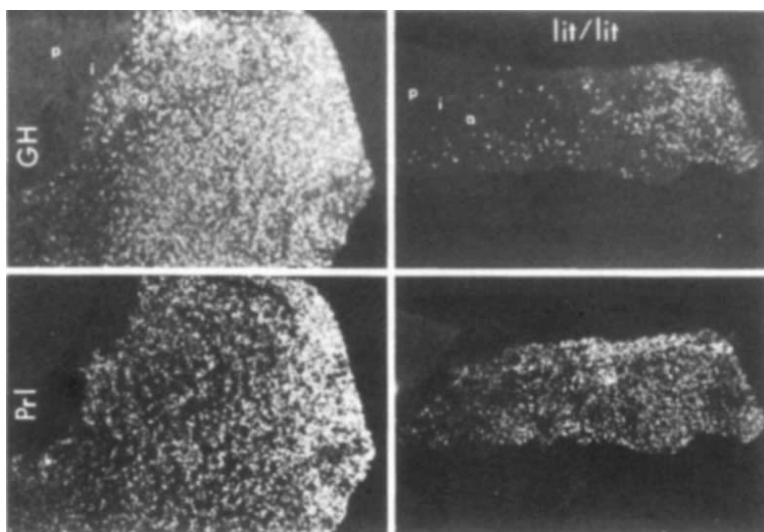
selective defect in caudomedial proliferation of somatotrophs. Photomicrographs of anterior pituitary glands from roughly 60-day-old *lit/lit* and wild-type mice subjected to *in situ* hybridization using specific GRF receptor, growth hormone, proopiomelanocortin, prolactin, and Pit-1  $^{35}S$ -cRNA probes. The *lit/lit* anterior pituitary gland (A) was consistently smaller (~twofold) than wild-type glands, strikingly in the lateral lobes. *In situ* hybridization consistently revealed no overall alteration in POMC-staining cells, but the signal was significantly compacted in the hypoplastic *lit/lit* anterior pituitary with comparable expression in the intermediate lobe (IL). Prolactin hybridization signals in *lit/lit* glands were minimally reduced compared to those in the wild-type glands and exhibited a normal uniform pattern across the entire pituitary gland. The pattern of Pit-1 hybridization, similar in *lit/lit* and wild-type mouse pituitaries, was observed to correspond to that of prolactin gene expression. In contrast, hybridization using growth hormone probe gave intense signals only in the anterolateral aspect of the gland and no detectable growth hormone-positive cells



were observed in the caudomedial portion of the gland. Similar data were reproducibly obtained in four individual *lit/lit* pituitary glands examined.

METHODS. The probes used in these experiments were as described<sup>4</sup> and exposed for 2 days for tissue slides hybridized with POMC, prolactin, growth hormone and 14 days with Pit-1, GRFR.

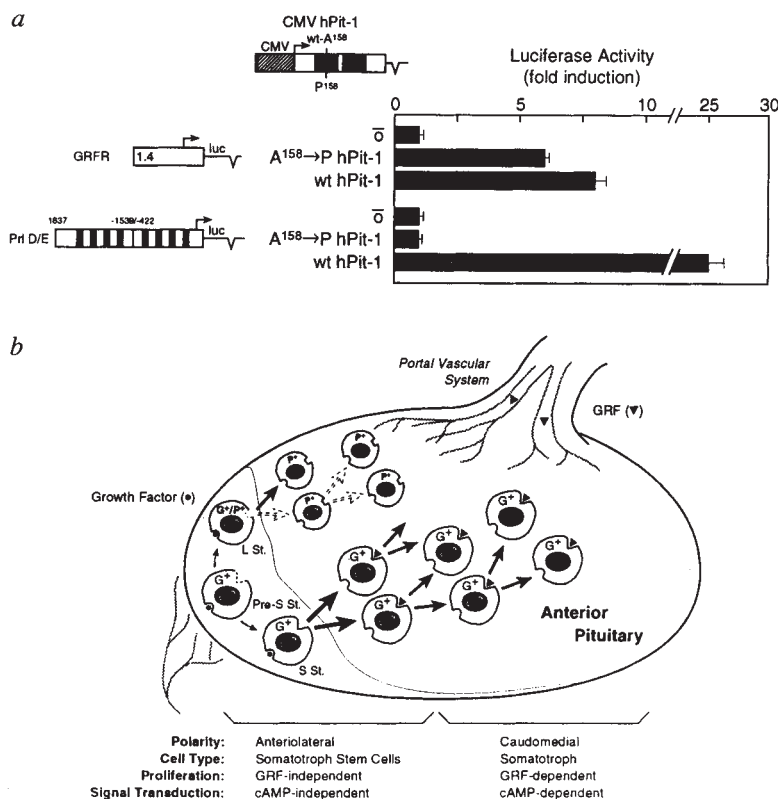
FIG. 5 Growth hormone (GH) and prolactin (PrI) immunoreactivities in wild-type (wt) and *little* (*lit/lit*) mouse pituitaries. Photomicrographs show individual frontal sections stained for growth hormone and prolactin using concurrent dual indirect immunofluorescence labelling. In each section, the positions of the posterior (p), intermediate (i) and anterior (a) lobes are indicated. The normal pituitary displays characteristically widespread and uniform distributions of cells stained for each hormone. The smaller and flattened *lit/lit* pituitary shows a nearly normal density of PrI cells, whereas detectable GH cells are markedly reduced in density and abnormally concentrated near the lateral margins of the gland. Very few cells revealed coexpression of growth hormone and prolactin. All magnifications,  $\times 75$ . METHODS. Mice were anaesthetized with chloral hydrate and perfused transcardially with neutral buffered 4% formalin. Pituitaries were postfixed overnight in the primary fixative with 10% sucrose added. Glands were frozen and sectioned at 20  $\mu\text{m}$ , and then exposed for 48 h at 4  $^{\circ}\text{C}$  to a mixture of monkey anti-mouse GH (1:80,000 dilution; provided by Y. Sinha, the Whittier Institute) and rabbit anti-human prolactin (Cambridge Medical Diagnostics, Billerica, MA). Primary antisera were localized using affinity-purified fluorescent-conjugated goat anti-mouse IgG (Cappel) and lissamine rhodamine-conjugated goat anti-rabbit IgG (Tago).



tinued replication of the cells, perhaps because they become physically removed from access to growth factors present in the stem cell mitotic zone (Fig. 6b). The portal venous system offers a distinct, restricted repertoire of growth factors delivered from the nerve termini in the median eminence of the hypothalamus. Besides proliferation roles, GRF could also exert functions in somatotroph survival, analogous to the dual survival and trophic roles of nerve growth factor (NGF). Thus, the mature anterior pituitary gland can be visualized as a stratified arrangement of two functional 'zones' of proliferation, each apparently under control of distinct growth factors.

In contrast to the profound defect in somatotroph proliferation, lactotroph proliferation in the *lit* mouse is apparently only mildly reduced compared to that found in the wild-type gland. The initial growth hormone-producing precursor of mature lactotrophs<sup>49</sup> must be effectively generated and differentiated in GRF receptor-defective mice, indicating that this receptor is not required for lactotroph stem cell proliferation. Although somatotroph cell division occurs across the entire gland, lactotrophs apparently emerge from the zone of stem cell proliferation as post-mitotic cells that are apparently 'pushed' to the centre of the gland.

FIG. 6 Pit-1-mediated GRF regulation of somatotroph proliferation. a, Regulation of the GRF receptor promoter by rat Pit-1 (rPit-1). Wild-type or W261→C Snell rat Pit-1 cDNAs under control of the CMV promoter<sup>5,12</sup> were cotransfected into CV1 cells with GRF receptor promoter/luciferase fusion genes or with a prolactin enhancer/promoter luciferase fusion gene. Results are average  $\pm$  s.e.m. of triplicate determinations. b, Model of GRF receptor modulation of somatotroph proliferation lineage. A presomatotroph stem cell (Pre-S St) is proposed to give rise to either a somatotroph-stem cell (S St) or lactotroph-stem cell (L-St) which at some point must express both growth hormone and prolactin ( $G^+$ ,  $P^+$ ). Somatotroph stem cells (S St) express growth hormone ( $G^+$ ); and at least a subset are GRF-receptor positive (V). Stem cells are located in the anterolateral circumference of the gland. Caudomedial proliferation of somatotrophs require GRF receptor and GRF, which is supplied by the portal circulation from median eminence. Somatotrophs proliferate (indicated by mitotic spindle) across the entire anterior pituitary. In contrast, lactotrophs, once differentiated, proliferate little, if at all (dotted arrow and solid nucleus). The two zones of proliferation, representing stem cells or mature somatotrophs, are cAMP-independent or cAMP-dependent, respectively. METHODS. Cotransfection analysis were done in exponentially growing HeLa cells (60 mm dishes) which were transfected with pCMV Pit-1 vectors (3  $\mu\text{g}$ ) and vectors expressing and GRFR promoter or prolactin enhancer/promoter-luciferase fusion genes (1  $\mu\text{g}$ ) using a calcium phosphate method<sup>5</sup>. After 48 h, luciferase activity was determined. The wild-type and Ala158→P human Pit-1 expression vectors were previously described<sup>50</sup>.





We suggest that many mammalian organs use similar regulatory strategies in controlling growth of specific, stratified cell populations. A critical advantage of a shift in growth factor usage is that distinct signal transduction pathways could regulate different sets of genes, simultaneously allowing both terminal differentiation and continued growth. In light of the severe

pituitary hypoplasia in the Pit-1-defective Snell mouse<sup>10</sup>, Pit-1 is likely to regulate other distinct growth factor receptors required for proliferation of thyrotroph and lactotroph stem cells. The *lit* mouse model adds a genetic basis to the suggestion that many of the vast group of 'gut-brain' peptides will exert critical roles as trophic factors. □

Received 6 May; accepted 14 June 1993.

1. Schwind, J. L. *Am. J. Anat.* **41**, 295–319 (1928).
2. Voss, J. M. & Rosenfeld, M. G. *Cell* **70**, 527–530 (1992).
3. Frawley, L. S. & Boockfor, F. R. *Endocr. Rev.* **12**, 337–355 (1991).
4. Simmons, D. N. *et al. Genes Dev.* **4**, 695–711 (1990).
5. Ingraham, H. A. *et al. Cell* **55**, 519–529 (1988).
6. Bodner, M. *et al. Cell* **55**, 505–518 (1988).
7. Mangalam, *et al. Genes Dev.* **3**, 946–958 (1989).
8. Fox, S. R. *et al. Molec. Endocr.* **4**, 1069–1080 (1990).
9. Dolle, P. *et al. Cell* **60**, 809–820 (1990).
10. Li, S. *et al. Nature* **347**, 528–533 (1990).
11. Snell, G. D. *Proc. natn. Acad. Sci. U.S.A.* **15**, 733–734 (1929).
12. Lin, C. R., Lin, S.-C., Chang, C. P. & Rosenfeld, M. G. *Nature* **360**, 765–768 (1992).
13. Wilson, D. B. & Wyatt, D. P. *Anat. Embryol.* **174**, 277–282 (1986).
14. Rouse, M., Bartke, A., Dumont, F. & Dubois, M. P. *Cell Tissue Res.* **223**, 415–420 (1982).
15. Eicher, E. M. & Beamer, W. G. *J. Hered.* **67**, 87–91 (1976).
16. Slabaugh, M. B., Lieberman, M. E., Rutledge, J. J. & Gorski, J. *Endocrinology* **109**, 1040–1046 (1981).
17. Jansson, J.-O., Downs, T. R., Beamer, W. G. & Frohman, L. A. *Science* **232**, 511–512 (1986).
18. Frohman, M. A., Downs, T. R., Chomczynski, P. & Frohman, L. A. *Molec. Endocr.* **3**, 1529–1536 (1989).
19. Cheng, T. C. *et al. Endocrinology* **113**, 1669–1678 (1983).
20. Landis, C. A. *et al. Nature* **340**, 692–696 (1989).
21. Rivier, J., Spiess, J., Thorner, M. & Vale, W. *Nature* **300**, 276–278 (1982).
22. Guillemin, R. *et al. Science* **218**, 585–587 (1982).
23. Thorner, M. O. *et al. J. clin. Invest.* **70**, 965–977 (1982).
24. Bilezikian, N., Swanson, L. W. & Vale, W. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6854–6857 (1986).
25. Burton, F. H., Hasel, K. W., Bloom, F. E. & Sutcliffe, J. G. *Nature* **350**, 74–77 (1991).
26. Seifert, H., Perrin, M., Rivier, J. & Vale, W. *Nature* **313**, 487–489 (1985).
27. Bilezikian, L. & Vale, W. *Endocrinology* **113**, 1726–1731 (1983).
28. Gick, G. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 1553–1555 (1984).
29. Mayo, K. E. *et al. Molec. Endocr.* **2**, 606–612 (1988).
30. Struthers, R. S., Vale, W. W., Arias, C., Sawchenko, P. E. & Montminy, M. R. *Nature* **350**, 622–624 (1991).
31. Ikeda, H. & Yoshimoto, T. *Cell Tissue Res.* **263**, 41–47 (1991).
32. Barinaga, M. *et al. Nature* **306**, 84–85 (1983).
33. Ambesi-Impiombato, F. S., Parks, L. A. M. & Coon, H. G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3455–3459 (1980).
34. Tsang, B. K., Rixon, R. H. & Whitfield, J. F. *J. Cell Physiol.* **102**, 19–26 (1980).
35. Roger, P. P., Servais, P. & Dumont, J. E. *FEBS Lett.* **157**, 323–329 (1983).
36. Levine, M. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 617–621 (1988).
37. Weinstein, L. S. *et al. N. Engl. J. Med.* **325**, 1688–1695 (1991).
38. Hen, R., Axel, R. & Obici, S. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4785–4788 (1989).
39. Mayo, K. E. *Molec. Endocr.* **6**, 1734–1744 (1992).
40. Gaylinn, B. D. *et al. Molec. Endocr.* **7**, 77–84 (1993).
41. Ishihara, T., Nakamura, S., Kazira, Y., Takahashi, T. & Nagata, S. *EMBO J.* **10**, 1635–1641 (1991).
42. Jappner, H. *et al. Science* **254**, 1024–1026 (1991).
43. Lin, H. Y. *et al. Science* **254**, 1022–1024 (1991).
44. Ishihara, T., Sigemoto, R., Mori, K., Takahashi, T. & Nagata, S. *Neuron* **8**, 815–819 (1992).
45. Thorens, B. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8641–8645 (1992).
46. Jelinek, L. J. *et al. Science* **259**, 1614–1616 (1993).
47. Behringer, R. R., Mathews, L. S., Palmiter, R. D. & Brinster, R. L. *Genes Dev.* **2**, 453–461 (1988).
48. Borrelli, E., Heyman, R. A., Arias, C., Sawchenko, P. E. & Evans, R. M. *Nature* **339**, 538–541 (1989).
49. Elliott, R. W. & Moore, K. J. *Mamm. Genome* **3**, S81–S103 (1992).
50. Pfäffle, R. W. *et al. Science* **257**, 1118–1121 (1992).
51. Carbajo-Perez, E. & Watanabe, Y. G. *Cell Tissue Res.* **261**, 333–338 (1990).
52. Green, E. L. *Genetics and Probability in Animal Breeding Experiments* (Oxford University Press, New York, 1981).
53. Warden, C. H., Fidler, J. S., Pace, M. J., Svenson, K. L. & Lulis, A. J. *J. clin. Invest.* (in the press).
54. Dietrich, W. *et al. Genetics* **3**, 425–447 (1992).
55. Lander, E. S. *et al. Genomics* **1**, 174–181 (1987).
56. Sanger, F., Micklen, J. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5466 (1979).

ACKNOWLEDGEMENTS. We thank Q. Zhang, Y. Xia and C. Arias for expert contributions; G. DiMattia for the A 158P hPit-1 vector and S. Rhodes and K. Lewis for discussions and comments; C. Nelson for his assistance with tissue culture; S. Martin for help in preparing this manuscript and C. Candolero and M. Schonneman for their assistance with figures. M.G.R. is an Investigator with the Howard Hughes Medical Institute. P.E.S. is an Investigator of the Foundation for Medical Research. This work was supported by NIH grants (to M.G.R., P.E.S. and A.J.L.).

## LETTERS TO NATURE

# A large nuclear accretion disk in the active galaxy NGC4261

Walter Jaffe\*, Holland C. Ford†‡, Laura Ferrarese†‡, Frank van den Bosch\* & Robert W. O'Connell§

\*Leiden Observatory, PB 9513, Leiden, 2300 RA, Netherlands

†Department of Astronomy and Physics, Johns Hopkins University, Baltimore, Maryland 21218, USA

‡Space Telescope Science Institute, 3700 St Martin Drive, Baltimore, Maryland 21218, USA

§Department of Astronomy, University of Virginia, Charlottesville, Virginia 22901, USA

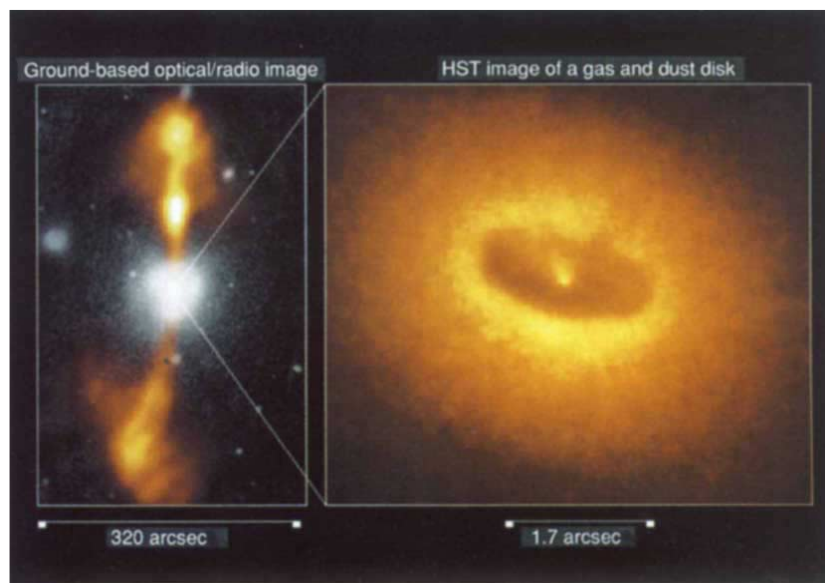
THE powerful emissions from the nuclei of active galaxies and quasars are thought to arise from the accretion of matter onto a massive black hole. Angular momentum will prevent matter from falling directly onto the central mass; instead, an 'accretion disk' should form, within which the gravitationally bound material will lose angular momentum and gradually spiral inwards. Accretion disks in active galactic nuclei have not hitherto been observed directly, but theoretical models<sup>1</sup> suggest that they are comparable in size to the gravitational radius of the central black hole (about  $10^{13}$  cm) and are heated to high temperatures by the energy released from the accreting matter. Using the Planetary Camera on the Hubble Space Telescope, we have discovered an unexpectedly large ( $2 \times 10^{20}$  cm radius) disk of cool dust and gas surrounding a bright unresolved nucleus in the

active galaxy NGC4261. We suggest that the bright point corresponds to thermal emission from the hot disk favoured by theoreticians, and that this is fuelled by material flowing from the cool 'outer' accretion disk. The spin axis of the accretion disk appears to determine the direction of the galaxy's radio jets.

Two exposures of 1,700 s and 700 s of NGC4261 were taken with the Planetary Camera (PC) on 15 January 1992 as part of a survey of a complete sample of elliptical galaxies in the Virgo cluster. The F550W filter used in both cases has an equivalent width of 1,200 Å centred at 5,500 Å. The two images were compared to identify and remove cosmic rays, then added with weights proportional to their exposure times. The image was deconvolved using the Richardson-Lucy algorithm<sup>2</sup> to correct for the blurring caused by the spherical aberration in the primary mirror of the Hubble Space Telescope (HST). The right-hand image of Fig. 1 displays a portion of the image containing the nucleus. The left-hand image, taken with the Palomar Mountain 48" Schmidt telescope, shows the galaxy at larger scale but with the same orientation. Superimposed in red is the 1,415 MHz radio continuum emission from the coincident radio source 3C270, imaged with the Very Large Array of the National Radio Astronomy Observatory (W. J. and B. R. McNamara, manuscript in preparation).

The emission in the outer areas of the HST picture arises from normal early type stars; their density increases smoothly towards the nucleus. Clearly visible in the centre of the image, however, is a sharply defined absorbing disk that blocks light from the stars behind it. At the very centre of the disk is a bright point-like nucleus. These features can be easily distinguished in

FIG. 1 Left: a composite image of NGC4261. In white is an optical image (V band) of the region around the galaxy taken by the Palomar Mountain 48" telescope, and in red the 1,415-MHz continuum radio emission, measured at the Very Large Array. East is down and north to the left. The numerous other white objects are either foreground stars or dwarf galaxies in the Virgo cluster. Right: a Hubble Space Telescope image, taken with the Planetary Camera, of the nuclear region of NGC4261 at  $\sim 100$  times greater magnification, but with the same orientation. The displayed region is  $4.25'' \times 3.75''$ ; individual pixels are  $0.44''$  square. The original monochromatic image has been deconvolved with the Lucy-Richardson algorithm and is displayed in a pseudocolour representation to emphasize subtle features. The 'temperature' pseudocolour representation used here shows the lowest intensities as black and dark red, intermediate intensities as orange and yellow, and the highest brightness as white.



the shaded surface representation of the data (Fig. 2). The disk itself is elliptical, with axes of  $1.71'' \times 0.74''$ , and its major axis is at a position angle (measured from north towards east) of  $-16^\circ$ . The position angle of the disk's major axis coincides with that of the galaxy as a whole<sup>3</sup> ( $-17^\circ$ ) but, surprisingly, the galaxy rotates about its major axis rather than its minor axis<sup>4</sup>. For an intrinsically circular disk the foreshortening implies that the plane of the disk is inclined  $64^\circ$  to the plane of the sky. We assume that the distance to NGC4261 is that of the Virgo cluster, 14.7 Mpc. In this case  $1'' = 71$  pc at the galaxy, or one PC pixel of  $0.044'' = 3.1$  pc. The projected disk size is then  $121$  pc  $\times$   $51$  pc. NGC4261 is in the "Virgo West" cloud (refs 6, 7; and T. Lauer, personal communication) and may be at about twice the distance of the main Virgo cluster; if this is true, the calculated physical dimensions of the disk would be doubled. The nucleus is unresolved in the PC image ( $< 0.1'' = 7$  pc) and has an apparent magnitude  $m_{5,500}$  of 23.6.

The absorption in the disk is presumably caused by interstellar dust embedded in cool gas. Dust absorption at the centre of NGC4261 was noted previously in ground-based pictures<sup>8,9</sup>, but the disk structure was not evident. A larger-scale dust feature (size  $\sim 20''$ ) has been described<sup>8</sup> but was not confirmed in our image. Our preliminary analysis shows that the face-on dust optical depth  $\tau$  of the disk is  $\sim 0.5$  (W. J., H. C. F., L. F., F. v. d. B. and R. W. O'C., manuscript in preparation). H I and CO radio spectra from the VLA and the James Clerk Maxwell Telescope (W. J. and B. R. McNamara, manuscript in preparation) and optical spectra from the William Herschel Telescope (W. J. *et al.*, manuscript in preparation), show that in addition to dust the disk contains atomic, molecular and ionized hydrogen, similar to dusty regions near ionizing sources in our Galaxy.

The radio source 3C270 has been mapped at various frequencies<sup>10</sup>, and consists of a nuclear point source and two jets with outer lobes. As can be seen in Fig. 1, the jet axis is parallel to the spin axis of the disk, and thus is perpendicular to the rotation axis of the galaxy. This strongly suggests that the angular momentum of the disk is responsible for the orientation of the jets. The effect might be caused either by transfer of angular momentum from the disk to a central black hole, or by formation of a jet 'nozzle' in the inner disk by hydrodynamic forces. A faint wedge-shaped emission feature extending  $\sim 0.4''$  west (left in Fig. 1) from the nucleus is visible in the HST image and may be related to the radio structures.

This image shows clearly the shape and size of the accretion disk of an active galactic nucleus (AGN), although its size and temperature differ greatly from those previously assumed. For

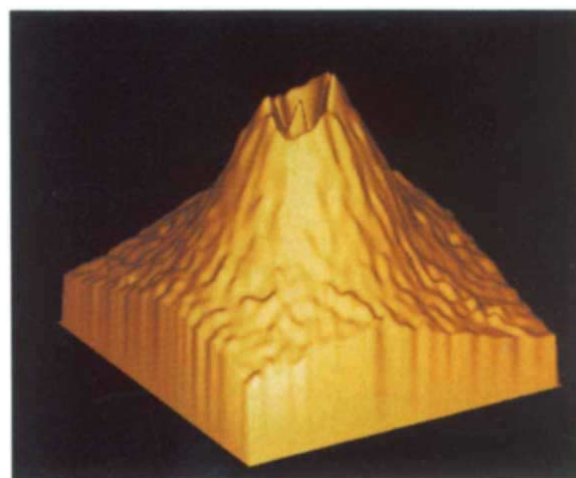


FIG. 2 The HST image displayed as a 'shaded surface' with intensity represented as the vertical axis seen in perspective. This emphasizes the drop in intensity at the disk and the bright inner peak. The image has been smoothed relative to Fig. 1 to reduce noise.

the supermassive black holes thought to be present in AGNs, the characteristic gravitational radius is a few astronomical units ( $1 \text{ AU} = 1.4 \times 10^{13} \text{ cm}$ ), and the disk rotation timescale is  $\sim 1,000$  s. Turbulent or magnetic viscosity acting in the spinning disk allows some material to lose angular momentum and fall towards the hole where the intense gravitational field accelerates, compresses and heats the material to a high temperature ( $\sim 10^6$  K). It radiates some of this energy away before falling through the gravitational radius, where it contributes to the mass and angular momentum of the hole. The energy released on this smallest size scale gives rise to such characteristics of AGNs as ultraviolet and X-ray emission, rapid variability and jet formation.

The disk that we observe in the HST picture is  $10^5$ – $10^7$  times larger than this presumed inner accretion disk, and the dynamic timescale at the outer rim, where the rotation speed is of the order of  $100 \text{ km s}^{-1}$ , is about  $10^6$  yr. Its temperature should be similar to that in dusty regions of our Galaxy, in the range 30 to  $10,000$  K. For disk turbulent velocities of the order of  $10 \text{ km s}^{-1}$ , we can estimate that the timescale for viscous transport<sup>11</sup> of 'fuel' into the central engine is about  $10^8$  yr. The time and energy



scales of the disk correspond to those of the large-scale radio jets and lobes. For example, the kinetic timescale in which the radio jet structures change is about  $10^6$  yr, while the synchrotron lifetimes of radio regions range from  $\sim 10^6$  yr for the hot spots to  $\sim 10^8$  yr for the more extended lobes. Similarly, the total 'equipartition' energy in the lobes,  $\sim 10^{57}$  erg, corresponds to the mass energy available in the disk: for a dust/gas ratio similar to our Galaxy, the disk mass is about  $10^5$  solar masses. Converted to energy at an efficiency of 1%, a value often assumed, this mass would yield  $10^{57}$  erg.

These facts, combined with the alignment of the radio axis and disk spin axis, lead us to describe the feature seen in NGC4261 as the 'outer accretion disk' of the central active nucleus. The bright unresolved point at the centre of the disk probably represents thermal optical emission from the hot inner accretion disk. The outer disk supplies fuel by way of the inner disk to the central engine, probably a massive black hole, in quantities that determine the luminosity, size and orientation of the extended radio emission.

Hints of such features have been obtained earlier: a previous ground-based image of NGC4261 showed a small central dust region ( $\sim 3''$  in diameter), but neither its size nor morphology could be accurately determined. Molecular radio observations<sup>12</sup> of Centaurus A have indicated the presence of rotating cold material in a rather larger region ( $\sim 1$  kpc). To our knowledge, however, the image presented here is the first of an accretion disk where size and structure can be directly associated with the results of nuclear activity.

On much larger scales, dust has been seen before in many other elliptical galaxies<sup>9,13</sup>, as lanes, disks and filamentary structures. These structures, often assumed to be the remnants of a captured late-type galaxy, are generally 10 to 100 times larger than the disk shown here. Although their presence may be correlated statistically with nuclear activity, their dynamic and decay timescales are much longer than those associated with AGN phenomena.

If the disk is in simple circular rotation, measurements of the rotation curve at HST resolution should lead to an estimate of the central mass that is free from the ambiguities of estimates derived from the orbits of stars. Measurements of the turbulent velocities should help to constrain models of the nature of the angular momentum and mass transport in the disk<sup>11</sup>.  $\square$

Received January; accepted May 1993.

1. Rees, M. J. *A. Rev. Astr. Astrophys.* **22**, 471–506 (1984).
2. Baade, D. & Lucy, L. B. *Messenger* **61**, 24–27 (1990).
3. Peletier, R. F., Davies, R. L., Illingworth, G. D., Davis, L. E. & Cawson, M. *Astr. J.* **100**, 1091–1142 (1989).
4. Davies, R. L. & Birkinshaw, M. *Astrophys. J.* **303**, L45–49 (1986).
5. Jacoby, G. H., Ciardullo, R., Ford, H. C. *Astrophys. J.* **356**, 332–349 (1990).
6. de Vaucouleurs, G. *Astrophys. J. suppl. Series* **6**, 213–234 (1961).
7. Binggeli, B., Sandage, A. & Tammann, G. A. *Astr. J.* **90**, 1681–1758 (1985).
8. Mollenhoff, C., Bender, R. *Astr. Astrophys.* **174**, 63–66 (1987).
9. Kormendy, J., Stauffer, J. in *IAU Symp. No. 127, 'Structure and Dynamics of Elliptical Galaxies'* (ed. de Zeeuw, P. T.) 405–406 (Reidel, Dordrecht, 1987).
10. Birkinshaw, M. & Davies, R. C. *Astrophys. J.* **291**, 32–44 (1985).
11. Pringle, J. E. *A. Rev. Astr. Astrophys.* **19**, 137–162 (1981).
12. Israel, F. P. et al. *Astr. Astrophys.* **227**, 342–350 (1990).
13. Bertola, F. in *IAU Symp. No. 127, 'Structure and Dynamics of Elliptical Galaxies'* (ed. de Zeeuw, P. T.) 135–143 (Reidel, Dordrecht, 1987).

## Century-scale effects of increased atmospheric CO<sub>2</sub> on the ocean–atmosphere system

Syukuro Manabe & Ronald J. Stouffer

Geophysical Fluid Dynamics Laboratory/NOAA Princeton University,  
PO Box 308, Princeton, New Jersey 08542, USA

SEVERAL studies have addressed the likely effects of CO<sub>2</sub>-induced climate change over the coming decades<sup>1–10</sup>, but the longer-term effects have received less attention. Yet these effects could be very significant, as persistent increases in global mean temperatures may ultimately influence the large-scale processes in the coupled ocean–atmosphere system that are thought to play a central part in determining global climate. The thermohaline circulation is one such process — Broecker has argued<sup>11</sup> that it may have undergone abrupt changes in response to rising temperatures and ice-sheet melting at the end of the last glacial period. Here we use a coupled ocean–atmosphere climate model to study the evolution of the world's climate over the next few centuries, driven by doubling and quadrupling of the concentration of atmospheric CO<sub>2</sub>. We find that the global mean surface air temperature increases by about 3.5 and 7 °C, respectively, over 500 years, and that sea-level rise owing to thermal expansion alone is about 1 and 2 m respectively (ice-sheet melting could make these values much larger). The thermal and dynamical structure of the oceans changes markedly in the quadrupled-CO<sub>2</sub> climate — in particular, the ocean settles into a new stable state in which the thermohaline circulation has ceased entirely and the thermocline deepens substantially. These changes prevent the ventilation of the deep ocean and could have a profound impact on the carbon cycle and biogeochemistry of the coupled system.

The model used here<sup>8</sup> consists of a general circulation model (GCM) of the atmosphere and oceans, and a simple model of land surfaces that includes the budgets of heat and water. It is a

global model with realistic geography. The atmospheric GCM includes the seasonal variation of insolation, and predicted cloud over which depends only on the relative humidity. It has nine vertical finite difference levels. The horizontal distribution of predicted variables is represented by spherical harmonics (15 associated Legendre functions for each of 15 Fourier components) and by corresponding grid-point values. The oceanic GCM uses a finite difference technique with a regular grid system which has horizontal spacing (4.5° latitude) by (3.75° longitude) and 12 vertical levels. This model is similar to that of Bryan and Lewis<sup>12</sup>, except that it mimics the effect of mesoscale eddies by the diffusion of potential temperature and salinity on isopycnal surfaces. The atmospheric and oceanic GCMs interact through the exchange of heat, water and momentum.

Assuming the temporal variations of atmospheric CO<sub>2</sub> in Fig. 1a, three 500-year integrations of the coupled model are done. One is a standard integration (S) in which the atmospheric CO<sub>2</sub> remains unchanged. In a second integration (4×C), the CO<sub>2</sub> concentration increases by 1% yr<sup>-1</sup> (compound) (close to the 'business as usual' (BAU) radiative forcing rate obtained by the Intergovernmental Panel on Climate Change<sup>13</sup>; IPCC) until it reaches four times the normal value at about the 140th year and remains unchanged thereafter. In a third integration (2×C), the CO<sub>2</sub> concentration also increases at the rate of 1% yr<sup>-1</sup> (compound) until it doubles around the 70th year and remains unchanged thereafter. By comparing the three integrations, one can evaluate the long-term impact of the doubling and quadrupling of atmospheric CO<sub>2</sub> on the coupled system.

The initial conditions for these integrations have realistic seasonal and geographical distributions of surface temperature, surface salinity and sea ice; the atmospheric and oceanic components of the model are nearly in equilibrium with these distributions. When the time integration of the model starts from this initial condition, the model climate rapidly drifts towards its own equilibrium state. To minimize the drift, the fluxes of heat and water at the ocean–atmosphere interface are adjusted by amounts that vary seasonally and geographically<sup>8</sup>. These adjustments, applied to all three integrations identified above, are independent of the anomalies of temperature and

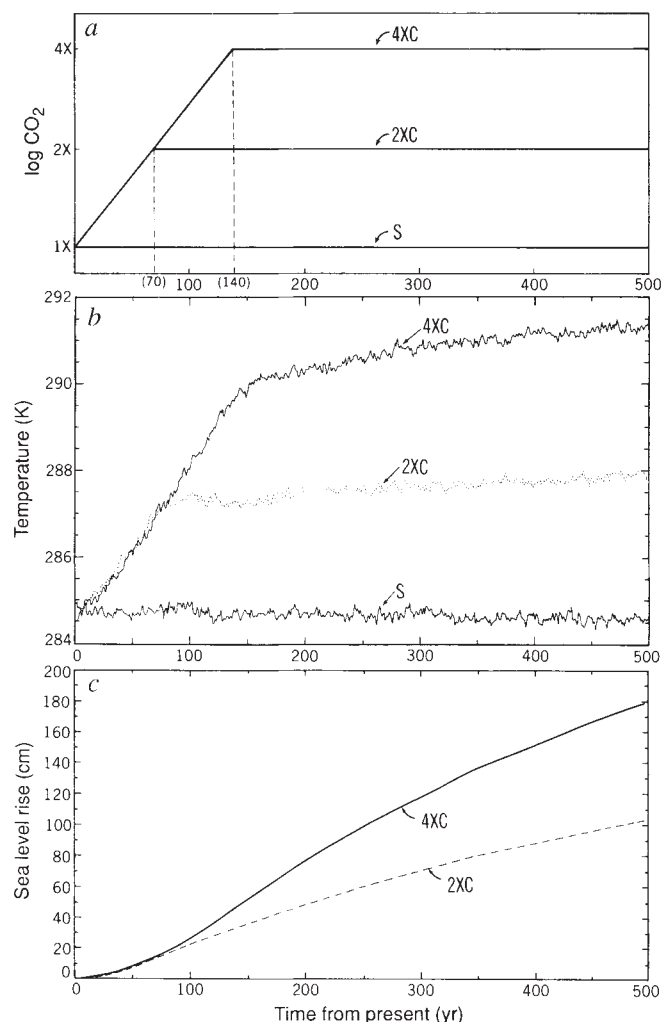


FIG. 1 Temporal variations of: *a*, logarithm of atmospheric CO<sub>2</sub> concentration; *b*, global mean surface air temperature (K); and *c*, global mean increase of sea level (cm) due to thermal expansion, computed as the difference between 4XC and S, and 2XC and S.

salinity at oceanic surface, and so neither damp nor amplify the anomalies.

Figure 1b contains the time series of global mean surface air temperature from the 4XC, 2XC, and S integrations. During the first 140 years of the 4XC integration, the global mean surface air temperature increases by 5 °C, at the rate of ~3.5 °C per century. After the 140th year, the global mean surface air temperature increases slowly by an additional 1.5 °C despite the absence of further CO<sub>2</sub> increase in the model atmosphere. The large thermal inertia of the deep ocean is mainly responsible for this residual warming.

A qualitatively similar feature is evident in the time series of the 2XC integration. During the first 70 years, the global mean surface air temperature increases by 2.2 °C, again at the rate of 3.5 °C per century. After atmospheric CO<sub>2</sub> stops increasing at the 70th year, the global mean surface air temperature increases by an additional 1 °C.

The temporal variations of global mean sea level due to thermal expansion of sea water alone are estimated for both the 4XC and 2XC integrations (Fig. 1c), although sea level is not explicitly predicted in the present model<sup>12</sup>. During the first few decades of the 4XC experiment, the sea level rises by ~1 cm per decade. The sea level continues to rise long after the 140th year when the atmospheric carbon dioxide stops increasing. A

qualitatively similar feature is indicated in the curve of sea-level rise in the 2XC integration. The total sea-level rise over the entire 500-year period of the 4XC amounts to about 1.8 m and is substantially larger than the corresponding rise of about 1 m in the 2XC.

Although the melt water from continental ice sheets is not included in the computation of sea-level rise mentioned above, the rate of melting at the surface of ice sheets has been estimated from the surface heat budget. If the effect of melt water were taken into consideration, the resulting sea-level rise could be much larger.

Figure 2 indicates that, in the 4XC, the thermohaline circulation (THC) almost disappears in most of the model oceans, leaving behind wide-driven cells. For example, the THC nearly vanishes in the North Atlantic during the first 200-yr integration (Fig. 3). In the immediate vicinity of the Antarctic continent, the THC weakens and becomes shallower (Fig. 2), markedly reducing the formation of Antarctic Bottom Water. This in turn weakens the northward flow of bottom water in both Pacific and Atlantic.

The near-extinction of the THC described above is attributable mainly to the capping of oceans by relatively fresh water in high latitudes, where the supply of water to the ocean surface increases markedly. The excess of precipitation over evaporation and runoff from continents increases in high-latitude oceans because of the enhanced poleward transport of water vapour in the warmer model troposphere.

The evolution of the THC in 4XC described above can be

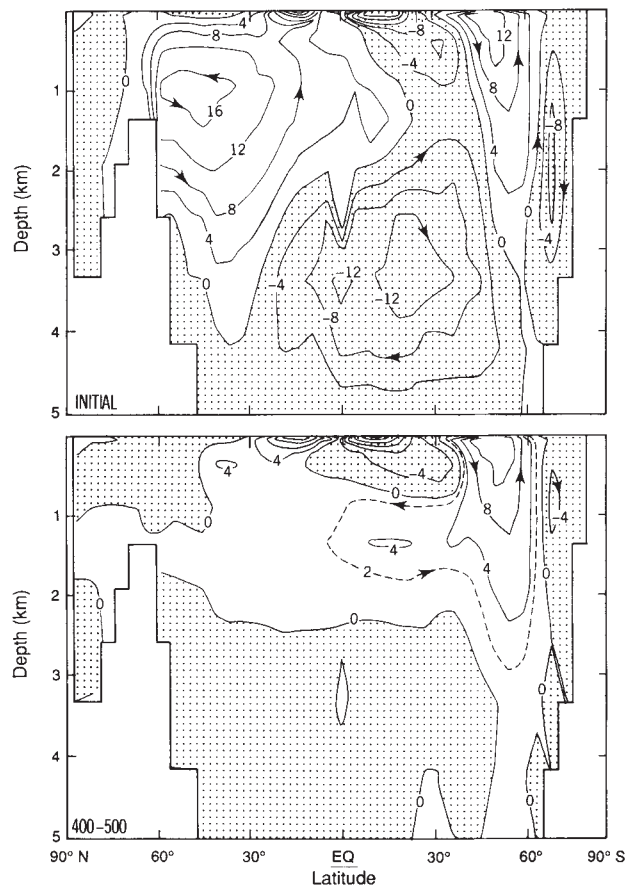


FIG. 2 Stream function of zonal mean meridional circulation in model oceans. Top: initial distribution obtained from the S. Bottom: average over the 400–500th year of the 4XC. Units on contours are in Sverdrups ( $10^6 \text{ m}^3 \text{ s}^{-1}$ ).



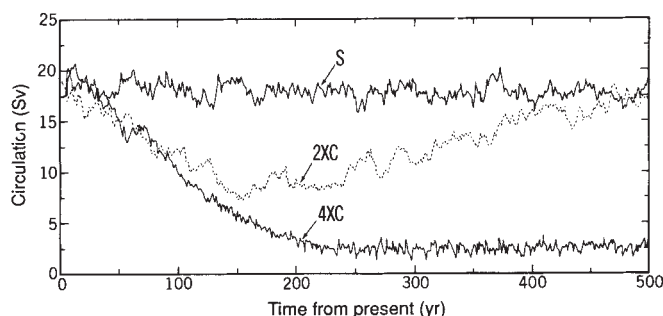


FIG. 3 Temporal variation of the intensity of the thermohaline circulation in the North Atlantic from the 4XC, 2XC and S. Here the intensity is defined as the maximum value of the stream function representing the meridional circulation (for example Fig. 2) in the North Atlantic.

compared with the 2XC in which the atmospheric carbon dioxide increases until it doubles (Fig. 3). In the 2XC, the intensity of the THC in North Atlantic again keeps decreasing long after the 70th year when atmospheric  $\text{CO}_2$  stops increasing and the rates of surface warming and freshening are abruptly reduced; it is eventually reduced to less than half of its original intensity. The THC begins to recover very slowly, however, around the 150th year of the integration. Because of the reduction of the upward advection of cold water that accompanies the weakening of the THC, a large warming of the subsurface layer takes place in the 2XC between the Equator and  $45^\circ\text{N}$  latitude in the North Atlantic. The resulting increase in the density contrast between the sinking and rising regions seems to be responsible for the gradual recovery of the THC in the North Atlantic by the 500th year.

The results from the 2XC described above imply that the weakening of the THC in the Atlantic during the early phases of both experiments does not represent the collapse of the circulation due to an instability. Instead, it represents a slow adjustment to the evolving density structure of the model Atlantic Ocean. Once the THC vanishes in the 4XC integration, however, it does not regenerate despite the intensification and deep penetration of positive temperature anomaly in low and middle latitudes. This suggests that the state of inactive THC reached in the 4XC is a stable state which is distinct from the state of the active circulation maintained in the early phase of both integrations. The existence of two stable equilibria in a coupled ocean-atmosphere model has been demonstrated previously<sup>14</sup>.

It is important to recognize that an adjustment of surface water flux is needed to sustain the THC in the Atlantic Ocean of the present model<sup>14</sup>. Although the adjustments of heat and water fluxes are independent of the anomalies of the sea surface temperature and salinity, and do not affect the damping rate of these anomalies, the sensitivity of the THC in the present

coupled model could be significantly different from the actual sensitivity. Therefore, even if atmospheric carbon dioxide is quadrupled, the near-disappearance of the THC may not occur in the real ocean. On the other hand, if the freshening of near-surface water due to the melting of continental ice sheets had been incorporated into the model, it could have induced the disappearance of the THC in the 2XC.

Towards the end of the 4XC experiment, the thermocline becomes deeper, particularly over the North Atlantic, where the near-disappearance of thermohaline circulation results in significant warming and increased salinity of intermediate water. The rise of surface air temperature (Fig. 4) is particularly large ( $11\text{--}16^\circ\text{C}$ ) over the Arctic Ocean where sea ice disappears almost completely during the warmer half of the year by the end of the  $\text{CO}_2$ -quadrupling integration. On the other hand, it is smaller ( $5\text{--}8^\circ\text{C}$ ) in the northern North Atlantic and the Circumpolar Ocean of the Southern Hemisphere. Here the increase of sea surface temperature is reduced because of vertical mixing of heat in the oceans<sup>8</sup>. The warming over continents ranges from  $7^\circ\text{C}$  to  $10^\circ\text{C}$  and is larger than that over oceanic regions with the exception of the Arctic Ocean mentioned above. The temperature change described above is almost as large as the difference between the present climate and the warm climate of the Late Cretaceous<sup>15</sup> approximately 65–90 million years ago. In the 2XC, the pattern of surface air temperature resembles, but is half as large as, the change in 4XC.

Averaged globally, the rate of increase in surface air temperature is  $\sim 3.5^\circ\text{C}$  per century during the first 150 years of the 4XC when the atmospheric concentration of carbon dioxide increases at the rate of  $1\% \text{ yr}^{-1}$  (roughly the BAU rate obtained by the IPCC). This is slightly larger than the IPCC's best estimate of  $3.0^\circ\text{C}$  per century. It has been estimated that the equilibrium response of the present model to the doubling of atmospheric  $\text{CO}_2$  is approximately  $3.5^\circ\text{C}$  in global mean surface air temperature. This response belongs in the upper half of the  $1.5\text{--}4.5^\circ\text{C}$  range of climate sensitivity estimated by the IPCC. Thus, it is possible that the present model may exaggerate the sensitivity of the actual climate.

According to the IPCC, a quadrupling of the  $\text{CO}_2$  equivalent of greenhouse gases could be realized if the BAU emission of greenhouse gases continued until the end of the twenty-first century. Draconian measures would probably be required to prevent the  $\text{CO}_2$  equivalent of greenhouse gases from quadrupling<sup>16</sup>. Considering the possible overestimate of climate sensitivity by the present model, it may be reasonable to speculate that the 2XC and 4XC experiments provide a probable range of future climate change. Nevertheless, one should not discard too readily the possibility of very large long-term climate change such as that which occurred in the 4XC experiment.

Received 4 May 1993; accepted 27 May 1993.

1. Bryan, K., Komro, F. G., Manabe, S. & Spelman, M. J. *Science* **215**, 56–58 (1982).
2. Schlesinger, M. E., Gates, W. L. & Han, Y. J. *Coupled Ocean-Atmosphere Models* (ed.

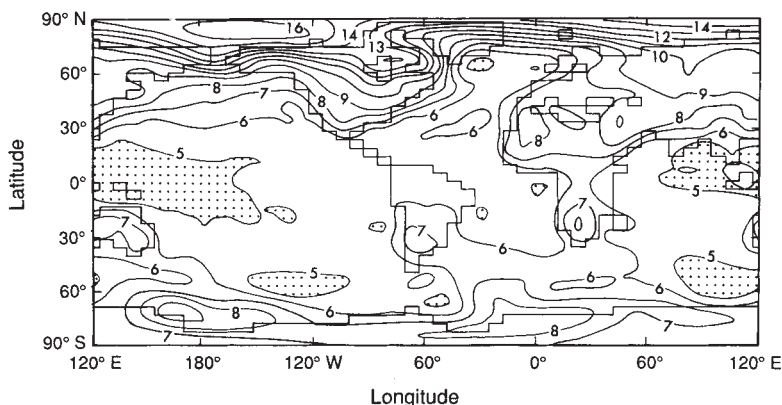


FIG. 4 Geographical distributions of the increase of surface air temperature from the  $\text{CO}_2$ -quadrupling experiment computed as the difference between the time mean values over the 400–500th year of the 4XC and the S integrations.

- Nihoul, J. C. J.) 447–478 (Elsevier, 1985).
3. Bryan, K. & Spelman, M. J. *J. geophys. Res.* **90**(C6), 11679–11688 (1985).
  4. Bryan, K., Manabe, S. & Spelman, M. J. *J. phys. Oceanogr.* **18**, 851–867 (1988).
  5. Washington, W. M. & Meehl, G. A. *Clim. Dynamics* **4**, 1–38 (1989).
  6. Stouffer, R. J., Manabe, S. & Bryan, K. *Nature* **342**, 660–662 (1989).
  7. Manabe, S., Bryan, K., Spelman, M. J. *J. phys. Oceanogr.* **20**, 772–749 (1990).
  8. Manabe, S., Stouffer, R. J., Spelman, M. J. & Bryan, K. *J. Clim.* **4**, 785–818 (1991).
  9. Hansen, J. *et al.* *J. geophys. Res.* **93**, 9341–9864 (1988).
  10. Cubasch, U. *et al.* *Clim. Dynamics* **8**, 55–69 (1992).
  11. Broecker, W. S. *Nature* **328**, 123–126 (1987).
  12. Bryan, K. & Lewis, L. J. *J. geophys. Res.* **84**, 2503–2517 (1979).

13. Intergovernmental Panel on Climate Change *Scientific Assessment of Climate*, WMO-UNEP (Cambridge Univ. Press, 1990).
14. Manabe, S. & Stouffer, R. J. *J. Clim.* **1**, 841–866 (1988).
15. Crowley, T. J. & North, G. N. *Paleoclimatology. Oxford Monogr. Geol. Geophys.* **16**, (1991).
16. Walker, J. C. G. & Kasting, J. F. *Paleogeogr. Paleoclimatol. Paleocool.* **97**, 151–189 (1992).

ACKNOWLEDGEMENTS. We thank K. Dixon for his advice on model improvements and J. D. Mahiman, K. Bryan and T. R. Toggweiler for their comments on the paper.

## Evidence for general instability of past climate from a 250-kyr ice-core record

W. Dansgaard\*, S. J. Johnsen\* †, H. B. Clausen\*, D. Dahl-Jensen\* N. S. Gundestrup\*, C. U. Hammer\*, C. S. Hvidberg\*, J. P. Steffensen\*, A. E. Sveinbjörnsdóttir †, J. Jouzel ‡ & G. Bond §

\* The Niels Bohr Institute, Department of Geophysics, University of Copenhagen, Haraldsgade 6, DK-2200 Copenhagen N, Denmark  
† Science Institute, Department of Geophysics, University of Iceland, Dunhaga 3, IS-107, Reykjavik, Iceland

‡ Laboratoire de Modélisation du Climat et de l'Environnement, CEA/DSM, CE Saclay 91191, and Laboratoire de Glaciologie et Géophysique de l'Environnement, BP 96, 38402 St Martin d'Hères Cedex, France

§ Lamont-Doherty Geological Observatory, Columbia University, Palisades, New York 10964, USA

RECENT results<sup>1,2</sup> from two ice cores drilled in central Greenland have revealed large, abrupt climate changes of at least regional extent during the late stages of the last glaciation, suggesting that climate in the North Atlantic region is able to reorganize itself rapidly, perhaps even within a few decades. Here we present a detailed stable-isotope record for the full length of the Greenland Ice-core Project Summit ice core, extending over the past 250 kyr according to a calculated timescale. We find that climate instability was not confined to the last glaciation, but appears also to have been marked during the last interglacial (as explored more fully in a companion paper<sup>3</sup>) and during the previous Saale–Holstein glacial cycle. This is in contrast with the extreme stability of the Holocene, suggesting that recent climate stability may be the exception rather than the rule. The last interglacial seems to have lasted longer than is implied by the deep-sea SPECMAP record<sup>4</sup>, in agreement with other land-based observations<sup>5,6</sup>. We suggest that climate instability in the early part of the last interglacial may have delayed the melting of the Saalean ice sheets in America and Eurasia, perhaps accounting for this discrepancy.

In 1990–92, the joint European Greenland Ice-core Project (GRIP) drilled an ice core to near the bedrock at the very top of the Greenland ice sheet (72.58° N, 37.64° W; 3,238 m above sea level<sup>7</sup>; annual mean air temperature –32 °C). The 3,028.8-m-long core was recovered by an electromechanical drill, ISTUK<sup>8</sup>. Less than 1 m of core in total was lost in the drilling process. The deepest 6 m is composed of silty ice with pebbles. The core quality is excellent, except for a 'brittle zone' in the depth interval between 800 and 1,300 m.

Here we discuss a timescale for the entire core and present a continuous profile of  $\delta^{18}\text{O}$  (hereafter denoted by  $\delta$ , the relative deviation of the  $^{18}\text{O}/^{16}\text{O}$  ratio in a sample from that in standard mean ocean water). In polar glacier ice,  $\delta$  is mainly determined

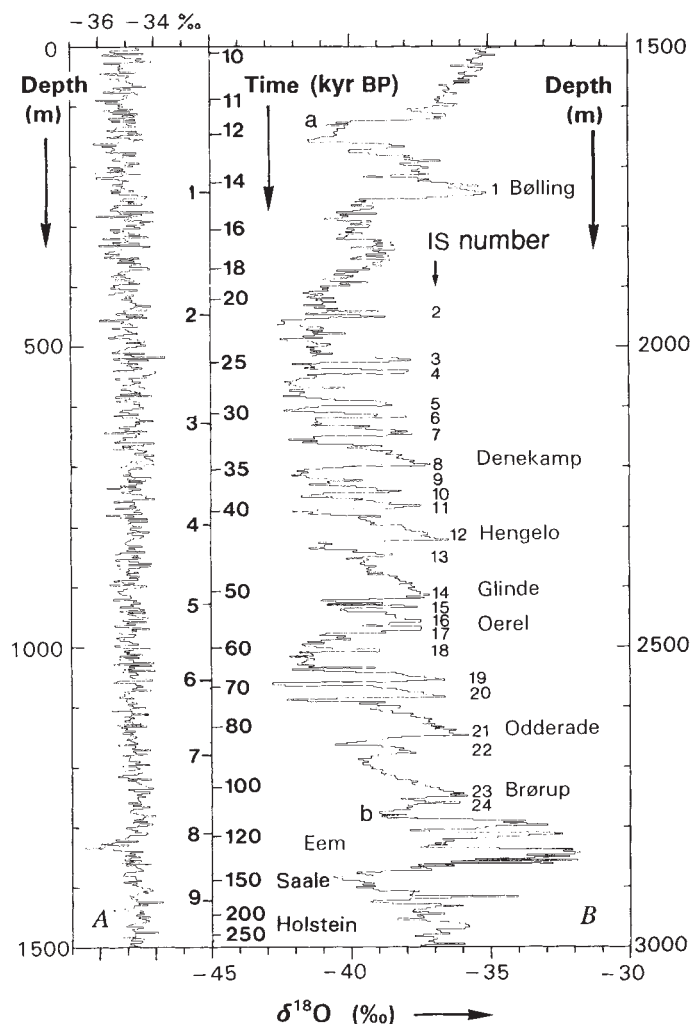


FIG. 1 The continuous GRIP Summit  $\delta^{18}\text{O}$  record plotted in two sections on a linear depth scale: A, from surface to 1,500 m; B, from 1,500 to 3,000 m depth. Each point represents 2.2 m of core increment. Glacial interstadials are numbered to the right of the B curve. The timescale in the middle is obtained by counting annual layers back to 14.5 kyr BP, and beyond that by ice flow modelling. The glacial interstadials of longest duration are reconciled with European pollen horizons<sup>16</sup>.

by its temperature of formation<sup>9</sup>. Profiles of several climate-related parameters spanning the last interglacial period (known as the Eemian period, or Eem) are being investigated by GRIP participants<sup>3</sup>.

The timescale back to 14.5 kyr BP (thousands of years before present) is derived by counting annual layers downward from



the surface<sup>1</sup>. Beyond 14.5 kyr BP a timescale is calculated as

$$t = \int_0^z dz / \lambda_z$$

$t$  being the age of the ice and  $\lambda_z$  the annual layer thickness at depth  $z$ . A steady-state ice flow model<sup>10</sup> is modified by introducing a sliding layer at bedrock<sup>11</sup>, and using a  $\delta$ -dependent accumulation rate ( $\lambda_H$ ) derived from all available  $\lambda_z$  data,

$$\lambda_H = \lambda_{H0} \exp [0.144 (\delta + 34.8)] \text{ m of ice equivalent per year,}$$

$\lambda_{H0}$  being the present value, 0.23 m yr<sup>-1</sup> (ref. 1).

The other flow model parameters are as follows: The total thickness of the ice sheet  $H = 3,003.8$  m of ice equivalent; the thickness of the intermediate shear layer  $h = 1,200$  m; the ratio between the strain rates at the top of the silty ice and at the surface,  $f_b = 0.15$ ; and the thickness of the silty ice layer,  $dh = 6$  m. The latter value may be too low, but higher values would significantly influence only the calculated ages of the deepest 50 m (ice older than 250 kyr) which will not be discussed here.

The  $h$  and  $f_b$  values are chosen so as to assign well-established ages to two characteristic features in the  $\delta$  record: 11.5 kyr for the end of the Younger Dryas event<sup>1,12</sup> and 110 kyr for the marine isotope stage (MIS) 5d<sup>4</sup>, which appear at depths of 1,624 m and 2,788 m, respectively, in the  $\delta$  record. These are points a and b in Fig. 1B. Back to 35 kyr BP the calculated timescale agrees essentially with that presented in ref. 1.

One of the assumptions behind the timescale calculation is that the stratigraphy has remained undisturbed, so that all annual layers are represented in a continuous sequence and thinned according to the depth-dependent vertical strain described by the flow model. This may fail at great depths if there is folding close to a hilly bedrock, and/or random thinning of layers of different rigidity (boudinage effect<sup>13</sup>).

Large-scale folding caused by bedrock obstacles hardly exists in the Summit area, however, because the bedrock is gently sloping ( $\leq 40$  m per km) in a large area around the drill site<sup>14</sup>. Furthermore, according to radio-echo sounding records (L. Hempel, personal communication) the shape of internal reflection layers suggests that the long-term position of the ice divide was only 5 km west of the present Summit. Consequently, the ice movement at Summit has been essentially vertical in the past, confirming that the ice cannot have travelled long distances over hilly bedrock. Nonetheless, at great depths the boudinage effect

may have caused small-scale disturbances. Some layers may have thickened at the expense of others now missing in the core. If so, the layer sequence may not be strictly continuous, but even then the broad outline of the timescale may still be valid.

Visible cloudy bands, probably indicative of former surfaces<sup>15</sup>, lie almost perpendicular to the core axis down to  $\sim 2,900$  m depth, corresponding to 160 kyr BP. Then follows a 54-m increment of apparently disturbed stratigraphy (S. Kipfstuhl, personal communication), possibly caused by the boudinage effect. Finally, the regular layer sequence is re-established from 2,954 m (210 kyr BP), but special caution should be applied beyond 160 kyr BP. The American GISP2 deep ice core being drilled 30 km west of Summit<sup>2</sup>, may provide verification of the timescale.

A continuous  $\delta$  record along the upper 3,000 m of the GRIP core is plotted on a linear depth scale in two sections of Fig. 1. The upper half of the record (Fig. 1A,  $\delta$  scale on top) spans nearly the past 10 kyr, and each  $\delta$  value represents the snow deposition through a few years near surface, increasing to 20 years at 10 kyr BP. Apart from the  $\delta$  minimum at  $8,210 \pm 30$  yr BP<sup>1</sup>, the record indicates a remarkably stable climate during the past 10 kyr.

In the contrast, the rest of the record shown in Fig. 1B is dominated by large and abrupt  $\delta$  shifts. Because of plastic thinning of the layers as they approach the bedrock, these 1,500 m of ice represent a much longer period of time than the 1,500 m above. On the right side of the record is an extension of the numbering of glacial interstadials (IS) introduced previously<sup>1</sup>. Furthermore, a series of European pollen horizons<sup>16</sup> (<sup>14</sup>C dated back to 60 kyr BP) is reconciled with the longest lasting  $\delta$ -based interstadials. The <sup>14</sup>C datings are all in essential agreement with our timescale once recent corrections to the <sup>14</sup>C scale<sup>12</sup> have been considered.

Figure 2 is a composite of five different chronological records. Figure 2D shows the upper 2,982 m of the Summit  $\delta$  record plotted on the linear timescale in 200 yr increments. The vertical line is drawn for comparison with the Holocene mean  $\delta$  value,  $-35\text{‰}$ . On the right-hand side of this line is a division of the last glacial cycles in European terminology. The adopted timescale is further supported, first by the numerous common features (particularly during the last glaciation) between the smoothed version and the other four records in Fig. 2 (exemplified by

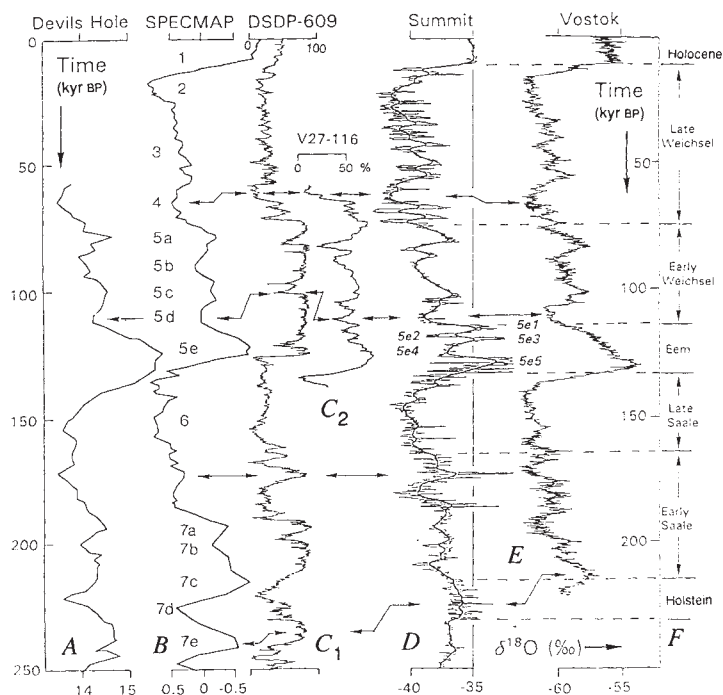


FIG. 2 Four climate records spanning the last glacial cycles and plotted on a common linear timescale. A,  $\delta^{18}\text{O}$  variations in vein calcite from Devils Hole, Nevada<sup>6</sup>. Dating by U/Th. B, The SPECMAP standard isotope curve<sup>4</sup> with conventional marine isotope stages and sub-stages. Dating by orbital tuning. C, part 1. Grey-scale measurements along 14.3 m of ocean sediment cores DSDP site 609. Part 2. Per cent  $\text{CaCO}_3$  in V27-116 (through stage 5) from locations west-southwest and west of Ireland. Scale in arbitrary units on top<sup>24</sup>. Dating by orbital tuning. D,  $\delta^{18}\text{O}$  record along the upper 2,982 m of the GRIP Summit ice core. Each point represents a 200-yr mean value. The heavy curve is smoothed by a 5-kyr gaussian low-pass filter. Dating by counting annual layers back to 14.5 kyr BP, and beyond that by ice flow modelling. Along the vertical line, which indicates the Holocene mean  $\delta$  value, is added an interpretation in European terminology. E,  $\delta\text{D}$  record from Vostok, East Antarctica<sup>5</sup>, converted into a  $\delta^{18}\text{O}$  record by the equation  $\delta\text{D} = 8 \cdot \delta^{18}\text{O} + 10\text{‰}$ . Dating by ice flow modelling.

dashed arrows); second, because a maximum entropy spectrum<sup>17</sup> comprising the total Summit record contains significant signals on the two Milankovitch cycles 41 kyr (obliquity) and, considerably weaker, 24/18 kyr (precession).

The violent  $\delta$  shifts observed in Greenland cores are less pronounced in the  $\delta$  record along the Vostok, East Antarctica, ice core<sup>5</sup> (Fig. 2E), probably because the shifts in Greenland are connected to rapid ocean/atmosphere circulation changes in the North Atlantic region<sup>18,19</sup>. Both records, as well as the  $\delta$  record from Devils Hole<sup>6</sup> (Fig. 2A), and the records in Fig. 2C introduced below, imply MIS 5e (Eem) as an interglacial of considerably longer duration than estimated from sea sediment  $\delta$  records, such as the SPECMAP record<sup>4</sup> in Fig. 2B. The disagreement may be explained, in part, by the climate instability recorded (Fig. 2D) in the early stages of Eem, which must have slowed the melting of the Saalean ice sheets in America and Eurasia. Further evidence of delayed sea level rise is found in records of the isotopic composition of atmospheric oxygen in the Vostok core<sup>5</sup>. If MIS 5e is defined as the period between the first and last years of higher than Holocene  $\delta$ , it lasted nearly 20 kyr, from 133 to 114 kyr BP, according to the Summit record.

In Fig. 2D (and in other Summit records<sup>3</sup>), 5e stands out as an interglacial abruptly interrupted several times by periods as cool as 5a and 5c. A similar episode in 5e was previously demonstrated in the Camp Century and Devon Island ice cores<sup>20,21</sup>, and given the new timescale, it could well be identical with one of the 5e cool spells indicated in Fig. 2D. The duration of 5e2 and 5e4 was apparently 2 and 6 kyr, and they define a tripartition of 5e. Carbon dioxide analyses along the Vostok ice core also suggest a partition of 5e<sup>22</sup>, but Fig. 2D differs remarkably from most pollen and deep sea records, which show the Eem as a generally warm and stable period (see for example, ref. 16 and Fig. 2B).

The apparent disagreement may not be serious. Substantial global climate changes are generally more pronounced at higher latitudes. The cooling in western Europe corresponding to the long-lasting  $\delta$  minimum 5e4 in Fig. 2D may therefore not have been deep enough to cause any marked change in the vegetation, and thereby in the pollen records. Sea sediment  $\delta$  records are primarily indicative of the continental ice volume, which does not necessarily vary with temperature during times of no ice in North America and Eurasia. Furthermore, the resolution is limited by bioturbation and coarse sampling. Significant fluctuations of 5e have been recorded, however, in North Atlantic areas of high sedimentation rate ( $\geq 5$  cm kyr<sup>-1</sup>), for example in core V28-56 from the Norwegian sea<sup>23</sup>, and as colour and CaCO<sub>3</sub> concentration changes in central North Atlantic sediment cores<sup>24</sup>.

The colour record from Deep Sea Drilling Project site 609 is plotted in Fig. 2C, part 1, on a timescale tuned<sup>25</sup> to orbital variations. This record has a resolution comparable to that of Fig. 2D. Disregarding MIS 2 and 3, where detrital grains of carbonate corrupt the grey scale, nearly all of the  $\delta$  shifts can be recognized in the colour record, including some of the  $\delta$  shifts in 5e. The colour scale reaches 'saturation' (note that 5a, and 5c and the warm phases of 5e appear equally dark), which may be why only one of the cool stages in 5e looks significantly different in colour. Several fluctuations in 5e are registered in cores recovered farther north<sup>24</sup>; for example see the inset of per cent CaCO<sub>3</sub> (indicative of biological activity) through MIS 5 from V27-116 in Fig. 2C, part 2.

The glacial cycles spanned by the Summit ice core appear different in Fig. 2D. For example, MIS 6 (late Saale) was apparently less cold and variable than late Weichsel, and a few spells of extreme warmth occurred in early Saale. Furthermore, the Holstein interglacial<sup>26</sup> (possibly MIS 7e) seems more stable than Eem, but less stable than Holocene.

In conclusion high resolution records suggest that, apart from the Holocene, instability has dominated the North Atlantic climate over the last 230,000 years. This applies to the Weichsel

glaciation (MIS 2 to 5d), to the Eem interglacial (MIS 5e) whose progress was very different from the Holocene, the Saale glaciation (MIS 6 to 7d), and Holstein, the preceding interglacial (MIS 7e), according to the interpretation given in Fig. 2D. This emphasizes the question of whether the Holocene will remain stable in spite of the growing atmospheric pollution. □

Received 4 April; accepted 3 June 1993.

1. Johnsen, S. J. *et al.* *Nature* **359**, 311–313 (1992).
2. Taylor, K. C. *et al.* *Nature* **361**, 432–436 (1993).
3. GRIP Members *Nature*, **364**, 203–207 (1993).
4. Martinson, D. G. *et al.* *Quat. Res.* **27**, 1–29 (1987).
5. Jouzel, J. *et al.* *Nature* (in the press).
6. Winograd, I. J. *et al.* *Science* **258**, 255–260 (1992).
7. Ekholm, S. & Keller, K. *Nat. Survey and Cadastre, Geodetic Div., Tech. Rep. No. 6* (Copenhagen, Denmark, 1993).
8. Gundestrup, N. S. & Johnsen, S. J. in *Greenland Ice Cores: Geophysics, Geochemistry and Environment* (eds Langway, C. C. Jr, Oeschger, H. & Dansgaard, W.) 19–22 (Am. Geophys. Un. Geophys. Monogr. **33**, Washington DC, 1985).
9. Johnsen, S. J., Dansgaard, W. & White, J. W. C. *Tellus* **B41**, 452–468, (1992).
10. Dansgaard, W. & Johnsen, S. J. *J. Glaciol.* **8**, 215–223 (1969).
11. Johnsen, S. J. & Dansgaard, W. in *The Last Deglaciation: Absolute and Radiocarbon Chronologies* (eds Bard, E. & Broecker, W. S.) (NATO ASI Series 2, 1992).
12. Bard, E. *et al.* *Radiocarbon* (in the press).
13. Staffelbach, T., Stauffer, B. & Oeschger, H. *Ann. Glaciol.* **10**, 167–170 (1988).
14. Hodge, S. M. *et al.* *J. Glaciol.* **36**, 17–27 (1990).
15. Hammer, C. U. *et al.* *J. Glaciol.* **20**, 3–26 (1978).
16. Behre, K.-E. & van der Plicht, J. *Veg. Hist. Archaeobot.* **1**, 111–117 (1992).
17. Ulyrch, T. J. & Bishop, T. N. *Rev. Geophys. Space Phys.* **13**, 183–200 (1975).
18. Oeschger, H. *et al.* in *Climate Processes and Climate Sensitivity* (ed. Hansen, F.) 299–306 (Am. Geophys. Un. Geophys. Monogr. 29, Washington DC, 1984).
19. Broecker, W. S., Bond, G., Klas, M., Bonani, G. & Wolfli, W. *Palaeoceanography* **5**, 469–477 (1990).
20. Dansgaard, W., Clausen, H. B., Johnsen, S. J. & Langway, C. C. *Quat. Res.* **2**, 396–398 (1972).
21. Paterson, W. S. B. *et al.* *Nature* **266**, 508–511 (1977).
22. Raynaud, D. *et al.* *Science* **259**, 926–934 (1993).
23. Kellogg, T. B. *Marine Micropaleontol.* **2**, 235–249 (1977).
24. Bond, G., Broecker, W., Lotti, R. S. & McManus, J. in *Start of a Glacial* (eds Kukla, G. J. & Went, E.) 185–205 (NATO ASI Series I 3, Springer, Heidelberg, 1992).
25. Ruddiman, W. F., Raymo, M. E., Martinson, D. G., Clement, B. M. & Bachman, J. *Palaeoceanogr.* **4**, 353–412 (1989).
26. Linke, G., Katzenberg, O. & Grün, R. *Quat. Sci. Rev.* **4**, 319–331 (1985).

ACKNOWLEDGEMENTS: We thank C. C. Langway and H. Tauber for reading the manuscript and making corrections and additions. This work is a contribution of the international Greenland ice-core Project (GRIP) organized by the European Science Foundation. We thank the GRIP participants and supporters for their cooperative effort. We also thank the funding agencies in Belgium, Denmark, France, Germany, Iceland, Italy, Switzerland and the United Kingdom, as well as the XII Directorate of CEC, the University of Iceland Research fund, the Carlsberg Foundation, and the Commission for Scientific Research in Greenland, for financial support.

## Decreased metal concentrations in ground water caused by controls of phosphate emissions

H. R. von Gunten\* & Ch. Lienert

Laboratorium für Radiochemie, Universität Bern, CH-3000 Bern-9, Switzerland

A NEW generation of phosphate-eliminating sewage treatment plants and the recent prohibition of phosphates in detergents have considerably reduced the concentrations of phosphate in surface waters<sup>1</sup>. But there has been concern that replacing phosphate in detergents with complexing agents might cause increased mobilization of heavy metals, and consequent pollution of ground waters<sup>2,3</sup>. Here we present a twelve-year analysis of phosphate, manganese and cadmium in the river Glatt, Switzerland, and in an adjacent aquifer which is infiltrated by the river water. Together with a reduction of phosphate concentrations in both river and ground water over the study period, we find lower groundwater concentrations of manganese and cadmium. We postulate that lower phosphate levels have decreased the amount of oxidizable organic carbon in the river bed, and hence have

\*Also at Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland.



created less reducing conditions in the infiltrating water. This in turn has resulted in decreased reductive dissolution of manganese and cadmium in the ground water. It thus appears that eliminating phosphate has led to an unexpected improvement in drinking water quality with respect to heavy metals.

In alpine and pre-alpine aquifers of quaternary glaciofluvial origin ground water is partly recharged by infiltrating rivers<sup>4</sup>. The impact of polluted surface waters on the quality of ground water (used as drinking water) is of major concern and it is very important to understand the bio-geochemical processes<sup>5</sup> occurring during infiltration. The river Glatt, and a well instrumented site at Glattfelden, Switzerland, have been extensively used to study these processes (see, for example, refs 6–11). These investigations have shown that temperature-controlled (micro-) biologically mediated processes in the river bed create seasonal cycles in the chemical composition of the infiltrating water. The amplitudes of these cycles have important implications for the ground water quality<sup>11</sup>.

The river Glatt (discharge rate  $\sim 8 \text{ m}^3 \text{ s}^{-1}$ ) originates in the small eutrophic lake Greifen and flows through the densely populated and industrialized area of Zurich. At the field site, polluted river water infiltrates steadily at a rate of  $\sim 0.5 \text{ m}^3$  per day per  $\text{m}^2$  through the sediments of the river bed into the water-saturated aquifer (gravel, sand, silt and clays) and becomes the uppermost layer of the local groundwater stream<sup>12</sup>. Between 1979 and 1992 water samples were taken monthly (or in longer, irregular intervals) from the river and from several observation wells at the experimental site. Submersible plastic pumps and  $0.45\text{-}\mu\text{m}$  in-line filters were used for sampling. Manganese, cadmium and other trace elements were determined in acidified samples either by atomic absorption spectroscopy or by inductively coupled plasma atomic emission spectroscopy. Anions were measured in separate samples by different methods, and oxygen and pH were determined using submersible electrodes (for details see ref. 10). The quality of the analyses (sampling, contamination, reproducibility and accuracy) was assessed by standards and blanks.

The river/ground water infiltration system is inherently very complex. The main bio-geochemical processes<sup>13</sup> occur in the sediments of the river bed<sup>11</sup>. They can be detected by changes in the water composition between the river and the ground water. The following processes are of importance for this work: 1) micro-biologically mediated decomposition (oxidation) of organic matter and biota is responsible for a reduction of oxygen and nitrate levels, and hence influences the redox potential of the infiltrating water<sup>11</sup>; 2) these reducing conditions enable the reduction and consequent dissolution of manganese (hydr)oxides and heavy metals from sediments of the river bed and from the aquifer material; 3) the higher water temperatures in summer enhance biological productivity and degradation of organic matter, and lead to maxima (for example P, Mn, Cd) and minima (including  $\text{O}_2$ ,  $\text{NO}_3^-$ ) in the groundwater concentrations of the observed products<sup>11</sup>. Organic matter, including aquatic biota and micro-organisms, is the essential component in these processes. The main source of organic matter is lake Greifen<sup>1,3</sup>. Sewage treatment plants and tributaries downstream of lake Greifen contribute only little to the organic load of the river<sup>1,3</sup>. During the present investigation the concentrations of dissolved organic carbon (DOC) and total organic carbon (TOC) remained constant in the river Glatt (refs 3, 7, 8, 14 and Ch. Koch, personal communication). Mean values of DOC and TOC over the years 1979–1991 were  $3.9 \pm 0.6 \text{ mg l}^{-1}$  and  $5.6 \pm 1.2 \text{ mg l}^{-1}$ , respectively. Organic matter is deposited into the sediments of the river bed, but is also the nutrient for aquatic biota.

The results of our investigations of temperature and concentrations of oxygen, nitrate, phosphate, manganese and cadmium are presented in Fig. 1 for groundwater samples from three wells, two at 5 m and one at 7 m from the river bank. The higher summer concentrations of phosphate in the ground water

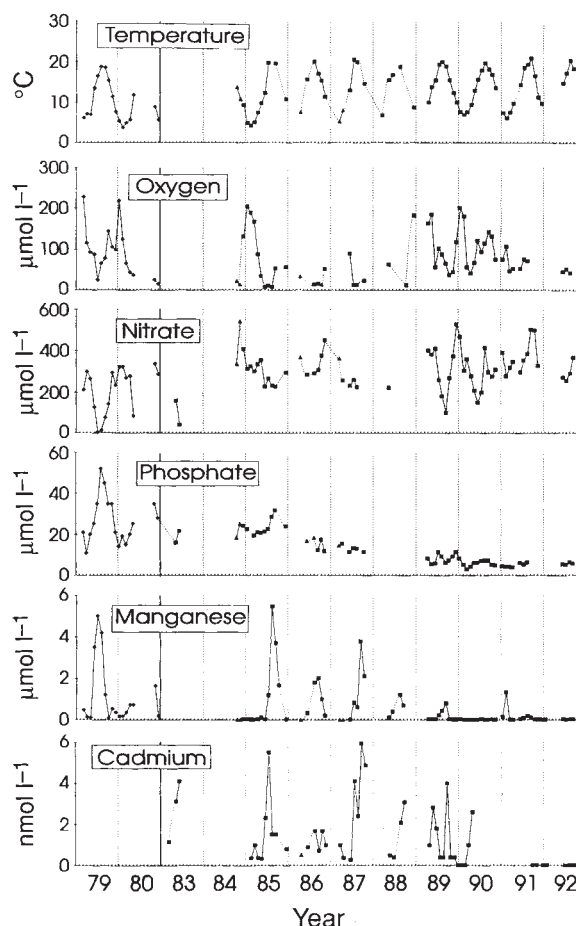
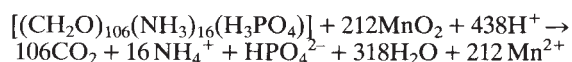


FIG. 1 Variations in 1979–80 and 1983–92 of temperature and dissolved oxygen, nitrate, phosphate, manganese and cadmium concentrations from three wells at Glattfelden, Switzerland. Two wells were at a distance of 5 m (◆, ■) and one at 7 m (▲) from the bank of the river. The samples were taken 50 cm below the groundwater table and passed through a  $0.45\text{-}\mu\text{m}$  filter. For analytical details see ref. 10. Data for 1979/80 were taken from refs 6–8 and J. Zobrist (personal communication), data for 1988 from M. Kuslys (personal communication) and other data from refs 9 (1983), 18 (1987) and 19–21.

resulted from the higher concentrations of phosphate in the river water<sup>1</sup> and from the decomposition of organic matter (for example aquatic biota, algae, humic substances) in the river bed. The higher concentrations of manganese in the summer were caused by a bacterial mediated reductive dissolution of manganese (hydr)oxides (see, for example, ref. 15) in the river bed and in the aquifer under the more reducing conditions (see equation below). The summer peaks of cadmium in the ground water resulted both from decomposed organic matter, and also the dissolution of manganese dioxide and calcite, which contain cadmium<sup>11,16</sup>. As an example of a possible redox process, we present an overall reaction (equation 1) for manganese dissolution involving organic matter (with a composition according to ref. 17).



It is obvious from Fig. 1 that since 1979 a significant and systematic decrease has occurred in the groundwater concentrations of phosphate, manganese and cadmium; the latter two elements have almost completely disappeared from the ground water after 1990. In contrast to these three species the average annual concentrations of oxygen and nitrate did not change (within errors). The significant decrease in groundwater phosphate concentrations resulted from the drastic reduction of

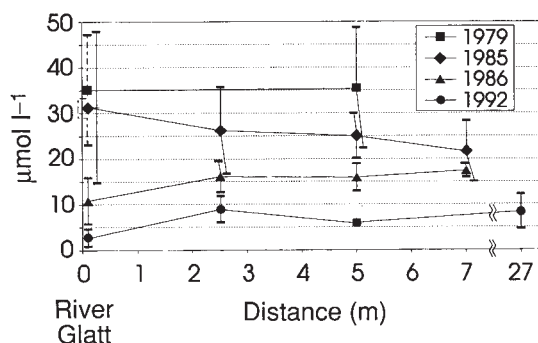


FIG. 2 Mean summer concentrations in 1979, 1985, 1986 and 1992 (50 cm below water table) of dissolved (0.45-µm filtered) phosphate in the river Glatt and in the groundwater samples from wells at 2.5, 5, 7 and 27 m from the river bank<sup>6-8,19-21</sup>.

phosphate in surface waters. This was due to improved, phosphate-eliminating sewage treatment plants and the prohibition of phosphates in detergents in 1986 (ref. 1).

Figure 2 shows the considerable changes in the mean summer concentrations of phosphate between 1979 and 1992 in the river Glatt, and in the groundwater sampling wells at distances of 2.5, 5, 7 and 27 m from the river bank. In this period we observed an approximately twenty-fold reduction in the mean summer concentrations of phosphate in the river water and a somewhat smaller decrease in the ground water. The phosphate concentrations in the ground water (that is values in the well, Table 1) resulted partly from decomposition of biota<sup>17</sup> and might be considered as a rough measure of productivity. It should be noted that after 1986 the concentrations of phosphate became significantly higher ( $5.4 \pm 1.9 \mu\text{mol l}^{-1}$ ) in the aquifer than in the river (Table 1), with a slightly decreasing trend. We postulate that the much smaller and still decreasing concentrations of phosphate in the river water have diminished productivity of aquatic biota in the river bed. The smaller amount of oxidizable biota in the river bed then led to locally less reducing conditions than existed before 1986, and hence to smaller concentrations of dissolved manganese and cadmium in the infiltrating water. At the much higher concentrations of oxygen and nitrate, (compared to manganese and cadmium) it is difficult to observe changes which are related to the better control of phosphate.

The concentrations of organic matter (DOC and TOC) in the river water remained constant during the twelve years of our investigation (ref. 14 and Ch. Koch, personal communication). Therefore, the supply of organic carbon cannot be the reason for a smaller productivity and the related decrease in the concentrations of manganese and cadmium, unless the quality of organic matter in the river has drastically changed due to improved sewage treatment techniques. Unfortunately only marginal data are available on this property.

TABLE 1 Phosphate and manganese concentrations in water ( $\mu\text{mol l}^{-1}$ )

| Year | River Glatt        |                  | Well at 5 m        |                  |
|------|--------------------|------------------|--------------------|------------------|
|      | $\text{PO}_4^{3-}$ | $\text{Mn}^{2+}$ | $\text{PO}_4^{3-}$ | $\text{Mn}^{2+}$ |
| 1979 | 35.1               | —                | 35.5               | 2.5              |
| 1985 | 31.3               | 0.06             | 24.9               | 2.1              |
| 1986 | 10.8               | 0.15             | 15.9               | 1.4              |
| 1987 | 3.8                | 0.21             | 12.7               | 1.3              |
| 1989 | 3.0                | 0.05             | 7.8                | 0.31             |
| 1990 | 1.3                | 0.10             | 6.9                | 0.02             |
| 1991 | 1.2                | 1.40*            | 6.2                | 0.12             |
| 1992 | 2.6                | 0.11             | 5.8                | 0.03             |

Values given are mean summer concentrations (May to September) in 0.45 µm filtered water. Note the high correlation of phosphate and manganese at 5 m distance from the river bank ( $R^2=0.91$ ).

\*Sampling of water with high particulate and colloidal contents after storm.

If the stoichiometry of equation (1) is approximately correct, the concentrations of decomposed organic matter (which determine the redox potential), phosphate and dissolved manganese should be correlated in the ground water at distances large enough to exclude the effects of local inhomogeneities of the river bed. Indeed, since 1979, we have found a strong correlation ( $R^2=0.91$ ) between the mean summer concentrations of dissolved phosphate and manganese (Table 1) in the ground water sampling wells at 5 m distance from the river bank. This is in contrast to river water, where no correlation between phosphate and the (relatively constant) concentration of manganese exists. Other parameters in the river water (for example the concentrations of cadmium, DOC and TOC) did not significantly change during the period of investigation.

We have demonstrated that before about 1990 microbial mediated bio-geochemical processes in the sediments of the river bed, involving the degradation of organic matter and aquatic biota, have created reducing conditions in the summer. The decreasing redox potential was then together with microbial activities, responsible for a release of heavy metals (such as manganese and cadmium) into the groundwater stream. The intensities of these processes have decreased unexpectedly, probably due to a more rigorous control of phosphate which was the only drastically changing nutrient in the period of investigation. Thus, the successful reduction of phosphate in surface waters not only fights eutrophication but has indirectly led to better groundwater quality with respect to harmful heavy metals. This result puts into perspective concerns about a possible remobilization of heavy metals (from aquifer material) by the partial replacement of phosphate in detergents by complexing agents. Also, we have not observed a significant change in the concentrations of nitrate since the prohibition of phosphates. These facts suggest that the measures taken by the Swiss authorities to improve water quality were appropriate. □

Received 18 January; accepted 1 June 1993.

- Davis, J. S. & Zobrist, J. *Mitteilungen EAWAG* **32**, 24–27 (1991).
- Giger, W. et al. *Mitteilungen EAWAG* **32**, 27–31 (1991).
- Mueller, B. & Sigg, L. *Aquat. Sci.* **52**, 75–92 (1990).
- Trüeb, E. *Gas-Wasser-Abwasser* **57**, 20–27 (1977).
- Sposito, G. *The Chemistry of Soils* (Oxford Univ. Press, 1989).
- Zobrist, J., Davis, J. S. & Hegli, H. R. *Gas-Wasser-Abwasser* **56**, 97–114 (1976).
- Hoehn, E., Zobrist, J. & Schwarzenbach, R. P. *Gas-Wasser-Abwasser* **63**, 401–410 (1983).
- Schwarzenbach, R. P., Giger, W., Hoehn, E. & Schneider, J. K. *Envir. Sci. Technol.* **17**, 472–479 (1983).
- von Gunten, H. R. & Kull, T. P. *Wat. Air & Soil Pollut.* **29**, 333–346 (1986).
- Jacobs, L. A., von Gunten, H. R., Keil, R. & Kuslys, M. *Geochim. cosmochim. Acta* **52**, 2693–2706 (1988).
- von Gunten, H. R. et al. *Geochim. cosmochim. Acta* **55**, 3597–3609 (1991).
- Hoehn, E. & Santschi, P. H. *Wat. Resour. Res.* **23**, 633–640 (1987).

- Stumm, W. & Morgan, J. J. *Aquatic Chemistry* (Wiley, New York, 1981).
- Hydrologisches Jahrbuch der Schweiz*, Landeshydrologie, Bern, Switzerland.
- Lovely, D. R. & Phillips, E. J. P. *Appl. envir. Microbiol.* **54**, 1472–1480 (1988).
- Davis, J. A., Fuller, C. C. & Cook, A. D. *Geochim. cosmochim. Acta* **51**, 1477–1490 (1987).
- Redfield, A. C. *Am. Scient.* **46**, 205–221 (1958).
- Waber, U. E., Lienert, Ch. & von Gunten, H. R. J. *Contaminant Hydrol.* **6**, 251–265 (1990).
- Kuslys, M. thesis, Univ. Bern (1988).
- Karametaxas, G. thesis, Univ. Bern (1991).
- Lienert, C. thesis, Univ. Bern (1992).

ACKNOWLEDGEMENTS. We thank M. Bradbury, H. Gäggeler, A. Grütter, E. Hoehn, U. Krähenbühl, S. Short, W. Stumm, U. von Gunten and J. Zobrist for discussions, and H. Ambühl, H. Bührer, Ch. Koch and J. Zobrist for unpublished data. We also thank G. Walker for constructive criticism and improvements to the manuscript, R. Keil for chemical analyses, and the Swiss National Science Foundation for partial support.



# Possibility of chemical weathering before the advent of vascular land plants

C. K. Keller\* & B. D. Wood†

\*Department of Geology, Washington State University, Pullman, Washington 99164-2812, USA

†Battelle, Pacific Northwest Laboratories, Richland, Washington 99352, USA

CHEMICAL weathering today is generally assumed to occur primarily in soils<sup>1,2</sup>. The rise of vascular plants during the Silurian and Devonian periods about 400 Myr ago brought about an increase in soil microbial activity and thus in soil CO<sub>2</sub> generation, and it has therefore been widely believed that, as a result of these changes, soil CO<sub>2</sub> replaced atmospheric CO<sub>2</sub> as the primary agent of chemical weathering<sup>3-6</sup>. Here we show that the aerated region above the water table (the vadose zone) exerts a strong influence on the CO<sub>2</sub> concentration to which runoff is exposed as it percolates beneath the soil, and we argue that this could have been the case before the Silurian. We present calculations which show that, for present-day atmospheric CO<sub>2</sub> concentrations, a low level of microbial respiration may be sufficient to support appreciable CO<sub>2</sub> concentrations in the vadose zone because of the slow rate of CO<sub>2</sub> loss to the surface. Despite the small amount of microbial respiration in pre-Silurian soils, CO<sub>2</sub> concentrations in subsoil vadose zones might therefore have been sufficient to account for the apparent constancy of chemical weathering since the mid-Proterozoic<sup>7</sup>, obviating the need to invoke high levels of atmospheric CO<sub>2</sub> to explain the weathering record.

For biogenesis of CO<sub>2</sub> to drive chemical weathering, there must be a respiring biological community and nutrients, particularly fixed (organic) carbon. Recent work suggests that respiration, and accompanying charging of percolating runoff with CO<sub>2</sub> and major elements, commonly occurs in subsoil vadose zones today. These processes are sustained wholly or in part by nutrients translocated from soils at the land surface<sup>8,9,43</sup>. Earlier data<sup>10-15</sup> are consistent with these processes. Recent *in situ* experiments<sup>16</sup> indicate that microbial respiration and silicate dissolution are enhanced in ground waters whose organic carbon levels are similar to those reported in vadose pore waters<sup>16,17</sup>, which is consistent with weathering in subsoil vadose zones. The question addressed here is the extent to which this could have occurred before the rise of higher land plants in the Silurian.

In a previous analysis applied to a different problem, Holland and Zbinden<sup>18</sup> neglected pre-Silurian soil and subsoil respiration and thereby effectively dismissed its weathering effect. It is true that early respiration rates should have been small relative to later rates, if only because of smaller rates of carbon fixation at Earth surface. However, the pre-Silurian terrestrial environment was not devoid of plants; the land surface probably sustained a microflora of algae and lichens<sup>19-21</sup> which would have generated a flux of organic matter available for rapid translocation into the subsurface through soils which were generally thin, sandy and rocky and thereby high in permeability and infiltration capacity<sup>21</sup>. Modern analogues might be the seasonal transport of high concentrations of dissolved organic matter to vadose zones and ground water through poorly-developed and/or microbially quiescent soils<sup>17,22,23</sup>. Such pulses can contain substantial fractions of organic acids<sup>17</sup> which presumably enhance carbonation weathering rates<sup>24</sup>; they could also support the microbial colonization of mineral grain surfaces, and the development of microreaction zones in which silicate dissolution has been observed to be concentrated in aquifer sediments<sup>16</sup>. The same could have been true of organic matter derived from primitive plant communities. Today, marine sedimentary organic carbon cycled into the shallow

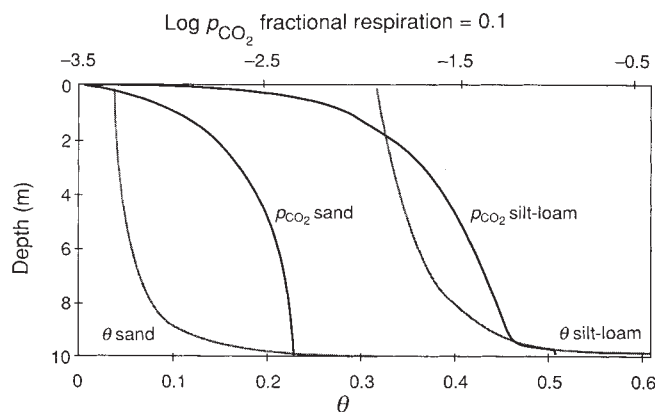


Fig. 1 Plots of two parameter values against depth in a hypothetical vadose zone. Shaded profiles show water contents  $\theta$  ( $L^3 L^{-3}$ ) calculated for hypothetical sand and silt-loam media. Assumed geometric mean and geometric standard deviation for the particle diameter are 0.7 mm and 5 respectively for the sand, and 0.03 mm and 7 respectively for the silt-loam. Assumed porosities are 0.4 (sand) and 0.52 (silt-loam). Solid profiles show simulated CO<sub>2</sub> concentrations for the two media, assuming total vadose respiration equals 0.1 times present-day soil respiration. Large  $p_{CO_2}$  values are attained at depth due to inhibition of CO<sub>2</sub> loss by low gas-phase diffusivity.

terrestrial subsurface seems to contribute in some cases to the substrate flux supporting subsurface microbial respiration<sup>8,15,23,25</sup>, and the possibility of autotrophic fixation of carbon in the subsurface cannot be excluded<sup>9</sup>. It must be assumed that these processes could also have augmented a land-surface-derived flux of fixed carbon in pre-Silurian time. And there is no reason to think that autotrophic and heterotrophic microbial communities have not existed in subsurface terrestrial sediments long before the Cambrian period<sup>26</sup>.

If conditions that existed which could have supported subsurface microbial respiration, albeit at small rates, then we need to assess the consequences for pre-Silurian subsurface CO<sub>2</sub> concentrations. For this analysis a steady-state, one-dimensional diffusion process following Fick's law is assumed and formulated as

$$\alpha = - \frac{\partial}{\partial z} \left[ D \frac{\partial C}{\partial z} \right]$$

where  $\alpha$  is CO<sub>2</sub> production rate (dimensions  $M L^{-3} T^{-1}$ ),  $z$  is depth beneath land surface (L),  $D$  is the effective gas-phase

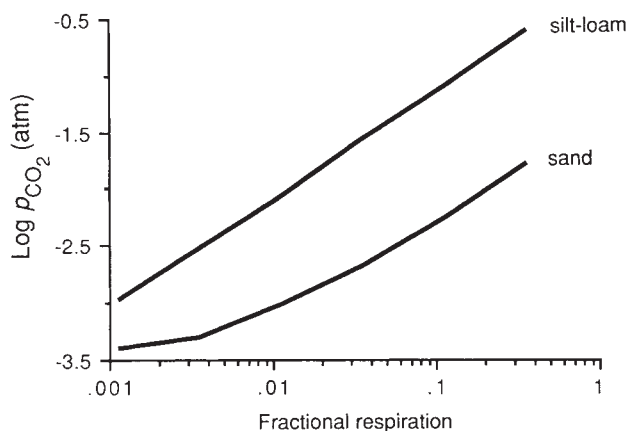
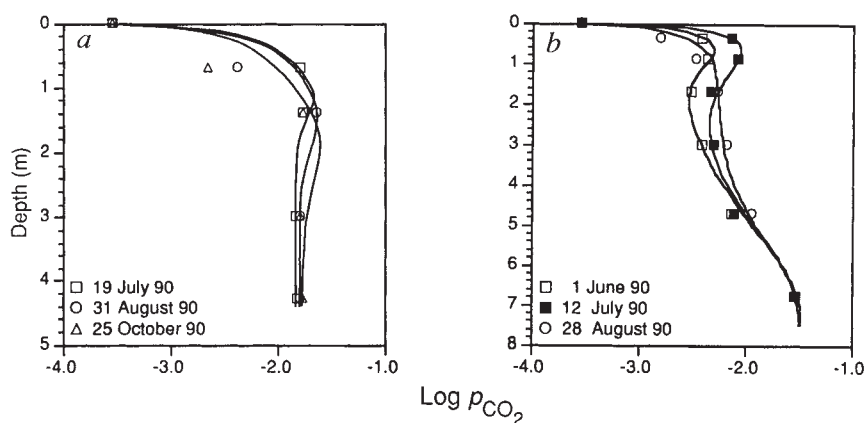


Fig. 2 Simulated  $p_{CO_2}$  at 9 m depth, plotted as a function of fraction of present-day soil respiration. Large  $p_{CO_2}$  values could be attained even at very small respiration rates for many media.

Fig. 3 Values of  $p_{\text{CO}_2}$  observed in vadose zones in (a) loess, southeastern Washington, USA, and (b) clayey till, southern Saskatchewan, Canada<sup>8</sup>. In both cases deep-vadose  $p_{\text{CO}_2}$  exceeds estimates of local mean soil  $p_{\text{CO}_2}$  by factors of 3 to 8. Calculated production rates at depth were similar to those assumed in the 0.1 fractional respiration simulation (Fig. 1).



diffusion coefficient in the porous medium ( $\text{L}^2 \text{T}^{-1}$ ), and  $C$  is concentration of  $\text{CO}_2$  in the gas phase ( $\text{ML}^{-3}$ ). Equation (1) was applied to a 10-m-deep model vadose zone to conform with the estimated global mean through geological time<sup>18</sup>. The boundary condition applied at the water table was zero-flux and the boundary condition at the land surface was specified concentration, set equal to present-day atmospheric concentration. These boundary conditions were kept invariant in order to study the effects of other factors as explained below. The governing equation (1) was discretized using a finite-difference scheme and solved iteratively for  $C$  at each depth node; spacing of the nodes was 0.1 m, and the accuracy of the numerical solution was checked by comparison with analytical solutions<sup>14,27</sup>. Details of the formulation and numerical methods are given by Wood *et al.*<sup>8</sup>.

The physics of this approach are similar to those used by Holland and Zbinden<sup>18</sup>, but in our formulation  $\alpha$  is not necessarily zero, and  $D$  distributions may be specified to vary with depth in the domain. The former capability permits modelling of non-zero respiration rates, and the latter accommodates the fact that  $D$  is not constant with depth in vadose zones, even in the case of a homogenous porous medium and uniform flow field. This is because water content, and hence the volume fraction occupied by interconnected gas phase through which gas-phase diffusion can occur, varies nonlinearly with matric potential ('tension' or 'suction') above the water table<sup>28</sup>. In nonindurated sediments, the range of this variation depends on grain size distribution. Coarse- and fine-grained cases were considered by calculating water contents  $\theta$  for grain size distributions typical of sand and silt-loam media, using the distribution statistics and methods of Campbell<sup>29</sup> for the simplest (hydrostatic) case (Fig. 1). Volumetric gas contents calculated from these water contents were used in relations provided in ref. 30 to calculate  $D$  for each node of model sand and silt-loam vadose zones.

Simulations were computed for cases in which  $\text{CO}_2$  production was distributed uniformly through the vadose zone. This distribution is consistent with rapid translocation and mixing of nutrients within the subsurface. The accumulative production flux, referred to in soils literature as 'soil respiration', was varied as a fraction of  $3.8 \times 10^{-5} \text{ g CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ , which is a conservatively small estimate of mean soil respiration in temperate grasslands today<sup>27,31,32</sup>. A fractional respiration of 0.001 is the approximate net primary productivity (NPP) of Antarctic rock surfaces, the smallest observed on Earth; fractional respiration of 0.01 corresponds to estimates of NPP in extreme desert, rock, sand and ice environments today<sup>33</sup>. The productivities of ancient lichen and algal communities may have been this small, but they certainly may have been larger, even as large as present-day values<sup>20,21,34</sup>.

Figure 1 shows simulated partial pressures of  $\text{CO}_2$  ( $p_{\text{CO}_2}$ ) plotted as a function of depth for the case in which fractional

respiration is 0.1. Most partial pressures of natural ground waters in today's temperate zones<sup>23,35,36</sup> appear to lie within the range bracketed by the two simulations near the water table. This result illustrates the fact that a small net respiration can yield large subsurface  $p_{\text{CO}_2}$  if part of that respiration occurs at depth in the vadose zone where small rates of diffusion slow the rate of  $\text{CO}_2$  loss to the surface<sup>8</sup>. The abrupt simulated increase of  $p_{\text{CO}_2}$  at the base of the silt-loam profile is associated with a zone of tension saturation just above the water table in which nearly all pore space is filled with water<sup>28</sup> and  $D$  therefore declines to very small values. Figure 2 shows how simulated  $p_{\text{CO}_2}$  at 9 m depth (above the zone of tension saturation) varies with fractional respiration. Although porous-medium characteristics exert strong control, even very small fractional respirations generally yield deep-vadose  $p_{\text{CO}_2}$  values that are several times the atmospheric level of  $10^{-3.5}$  bar. Only the smallest of fractional respirations shown, in the coarser medium, yield  $p_{\text{CO}_2}$  values indistinguishable from the upper atmospheric boundary. No percolation of ground water was considered in these simulations, and therefore the  $\text{CO}_2$  sink of dissolution into an advecting aqueous phase is not accounted for. Calculations assuming a range of recharge rates and using equilibrium relations provided by Wood *et al.*<sup>8</sup> indicate that the dissolution sink would have negligible effects on vadose  $p_{\text{CO}_2}$  at fractional respiration values down to 0.01. Calculations of the long-term bicarbonate flux from the continents by Holland and Zbinden<sup>5</sup> support this conclusion.

The profiles in Fig. 1 resemble the few present-day profiles available<sup>8,11,12,14</sup> (see also Fig. 3), showing that the model can produce subsurface weathering conditions similar to today's. The shapes are also consistent with the observed concentration of base-cation depletion by weathering in the shallowest parts of modern and ancient soils, where  $p_{\text{CO}_2}$  increases most rapidly along the percolation path, and with observed declines in intensity of base-cation depletion with increasing depth in soil-vadose continua. Integrated with respect to total vadose-zone depth, these deep depletions probably contribute substantially to the total weathering flux in many settings. One example<sup>37</sup> indicated that base-cation depletion in the upper 1 m of a thick till deposit may only be one-third to one-fourth of total base-cation depletion from the till due to weathering. In many settings such diffuse deep depletion would not be detectable by conventional means of sampling and analysis<sup>38</sup>.

The implication of the modelling results is that small vadose respiration rates could have charged early terrestrial runoff with carbonic acid and weathered mineral mass to levels similar to post-Silurian levels, even if atmospheric  $p_{\text{CO}_2}$  values were as small as today's. Although this hypothesis is very difficult to validate, it is consistent with the apparent continuity of terrestrial chemical weathering rates across the lower Palaeozoic soil development threshold. The large scatter of available  $\delta^{13}\text{C}$  observations for pre-Jurassic pedogenic carbonate<sup>39</sup> is also



consistent with our hypothesis: applying the steady-state diffusion model of Cerling<sup>27,40</sup> to our range of media properties and respiration rates, without varying atmospheric  $p_{\text{CO}_2}$  or  $\delta^{13}\text{C}$ , yields pedogenic carbonate  $\delta^{13}\text{C}$  values ranging from  $-12.5$  to  $+5.0$ ‰ at 1 m depth. Most respiration–diffusion combinations that we considered yielded results consistent with available data<sup>39,41</sup> for pre-Jurassic pedogenic carbonates.

One might wonder whether low oxygen concentrations in pre-Silurian soils, relative to those of today, might have suppressed microbial activity. We have shown previously, however, that although increases in  $p_{\text{CO}_2}$  in the vadose zone are roughly stoichiometric with decreases in  $p_{\text{O}_2}$  (indicating aerobic respiration), in neither of the settings that we studied did  $p_{\text{O}_2}$  approach the lower threshold for aerobic respiration. Thus it appears that atmospheric  $p_{\text{O}_2}$  could be substantially smaller than at present before it limited respiration in the vadose zone.

A pre-Silurian vadose weathering agency is thus theoretically plausible, and there is some evidence that it has present-day analogues. The argument is similar to that of Schwartzman and Volk<sup>24</sup>, who suggested that primitive terrestrial organisms may have substantially enhanced pre-Silurian weathering. Neither this conception nor recent arguments against it<sup>42</sup> accounts, however, for the transport and fate of organic acids and other microbial respiration substrates beneath rock surfaces and soils. We emphasize that these processes could affect chemical weathering, and the need for laboratory and field experimentation to understand the processes better. □

Received 5 April; accepted 25 May 1993.

- Nahon, D. B. *Introduction to the Petrology of Soils and Chemical Weathering* (Wiley, 1991).
- Drever, J. I. *The Geochemistry of Natural Waters* (Prentice-Hall, Englewood Cliffs, 1988).
- Cawley, J. L., Burruss, R. C. & Holland, H. D. *Science* **165**, 391–392 (1969).
- Holland, H. D., Lazar, B. & McCaffrey, M. *Nature* **320**, 27–33 (1986).
- Berner, R. A. *Science* **249**, 1382–1386 (1990).
- Berner, R. A. *Am. J. Sci.* **291**, 339–376 (1991).
- Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* (Princeton Univ. Press, 1984).
- Wood, B. D., Keller, C. K. & Johnstone, D. L. *Water Resources Res.* **29**, 647–659 (1993).
- Severson, K. J., Johnstone, D. L., Keller, C. K. & Wood, B. D., *Geomicrobiol. J.* **9**, 197–216 (1992).
- Kunkler, J. L. *U.S. Geol. Surv. Prof. Pap.* 650-B, 185–188 (1969).
- Reardon, E. J., Allison, G. B. & Fritz, P. *J. Hydrol.* **43**, 355–371 (1979).
- Haas, H., Fisher, D. W., Thorstenson, D. C. & Weeks, E. P. *Radiocarbon* **25**, 301–314 (1983).
- Thorstenson, D. C., Weeks, E. P., Haas, H. & Fisher, D. W. *Radiocarbon* **25**, 315–346 (1983).
- Wood, W. W. & Petraitis, M. J. *Water Resources Res.* **20**, 1193–1208 (1984).
- Keller, C. K. & Hendry, M. J. *Eos* **69**, 1178–1179 (1988).
- Hiebert, F. K. & Bennett, P. C. *Science* **258**, 278–281 (1992).
- Antweiler, R. C. & Drever, J. I. *Geochim. cosmochim. Acta* **47**, 623–629 (1983).
- Holland, H. D. & Zbinden, E. A. in *Physical and Chemical Weathering in Geochemical Cycles* (eds Lerman, A. & Meybeck, M.) 61–82 (Kluwer, The Netherlands, 1988).
- Schopf, J. W. *Biol. Rev.* **45**, 319–352 (1970).
- Golubic, S. & Campbell, S. E. *Precamb. Res.* **8**, 201–217 (1979).
- Beerbower, R. in *Geological Factors and the Evolution of Plants* (ed. Tiffney, B. H.), 47–91 (Yale Univ. Press, 1985).
- Thurman, E. M. *Organic Geochemistry of Natural Waters* (Martinus Nijhoff, The Netherlands, 1985).
- Keller, C. K. *Water Resources Res.* **27**, 2555–2564 (1991).
- Schwartzman, D. W. & Volk, T. *Nature* **340**, 457–460 (1989).
- Murphy, E. M., Davis, S. N., Long, A., Donahue, D. & Timothy Jull, A. J. *Water Resources Res.* **25**, 1893–1905 (1989).
- Ghiorse, W. C. & Wilson, J. T. *Adv. appl. Microbiol.* **33**, 107–172 (Academic, New York, 1988).
- Cerling, T. E., *Earth planet. Sci. Lett.* **71**, 229–240 (1984).
- Hillel, D. *Fundamentals of Soil Physics* (Academic, 1980).
- Campbell, G. S. *Soil Physics with BASIC: Transport Models for Soil–Plant Systems* (Elsevier, 1985).
- Millington, R. J. & Quirk, J. P. *Trans. Faraday Soc.* **57**, 1200–1207 (1961).
- Schlesinger, W. H. *Biogeochemistry: an Analysis of Global Change* (Academic, New York, 1991).
- Whittaker, R. H. *Communities and Ecosystems* (Macmillan, 1975).
- Johnson, C. G. & Vestal, J. R. *Appl. Envir. Microbiol.* **57**, 2308–2311 (1991).
- Knauth, P. L. *Geol. Soc. Am. Abstr. Prog.* **24**, A99 (1992).
- White, C. E., Hem, J. D. & Waring, G. A., in *Data of Geochemistry* 6th Edn (ed. Fleischer, M.) (U.S. Geol. Surv. Prof. Pap. 440-F, 1963).
- Harmon, R. S., White, W. B., Drake, J. J. & Hess, J. W. *Water Resources Res.* **11**, 963–967 (1975).
- April, R., Newton, R. & Coles, L. T. *Geol. Soc. Am. Bull.* **97**, 1232–1238 (1986).
- Keller, C. K., van der Kamp, G. & Cherry, J. A. *Water Resources Res.* **27**, 2543–2554 (1991).
- Mora, C. I., Dreise, S. G. & Seager, P. G. *Geology* **19**, 1017–1020 (1991).
- Cerling, T. E. *Am. J. Sci.* **291**, 377–400 (1991).
- Yapp, C. J. & Poets, H. *Nature* **355**, 342–344 (1992).
- Cochran, M. F. & Berner, R. A. in *Water–Rock Interaction 7* (eds Kharaka, Y. K. & Maest, A. S.) 473–476 (Balkema, Rotterdam, 1992).
- Hendry, M. J., Lawrence, J. R., Zanyk, B. N., & Kirkland, R. *Water Resources Res.* **29**, 973–984 (1993).

ACKNOWLEDGEMENTS. We thank P.E. Rosenberg, R. O'Brien and T.E. Cerling for their comments on the manuscript.

## Global correlations of mid-ocean-ridge basalt chemistry with seismic tomographic images

Eric Humler\*†, Jean Louis Thiro†  
& Jean Paul Montagner‡

\* Laboratoire de Géochimie et de Cosmochimie, Institut de Physique du Globe, 4 Place Jussieu, 75252 Paris Cedex 05, France

† Lamont Doherty Geological Laboratory, Palisades, New York 10964, USA

‡ Laboratoire de Séismologie, Institut de Physique du Globe, 4 Place Jussieu, 75252 Paris Cedex 05, France

THERMAL convection in the mantle is the most important dynamic process in the Earth's interior. Two quite different approaches are currently used to provide and analyse information regarding lateral temperature variations in the upper mantle. On the one hand, the major-element chemistry of basalts erupted at mid-ocean ridges is directly influenced by the temperature of the mantle beneath<sup>1–5</sup>; on the other, seismic velocity studies<sup>6–17</sup> can be used to map lateral temperature variations on a global scale (lower seismic velocities corresponding to higher mantle temperatures). Here we make a first attempt at relating these two approaches, by examining the global co-variation between mid-ocean-ridge basalt chemistry and upper-mantle shear-wave velocity. We find a strong correlation between basalt chemistry and variations in seismic velocity at depths of 100–170 km and lateral scale-lengths of 1,000–2,400 km, supporting a common thermal origin for the two types of signal. The departure of a few points from the general correlation may either reflect locally poor resolution of the tomographic model or, more interestingly, point to real anomalies that will repay closer examination.

The major-element data used to characterize mantle processes contain dispersion introduced by factors such as low- and high-pressure fractionation, partial melting and major-element heterogeneities in the source. Combining theoretical descriptions of mantle melting with experimental studies, Klein and Langmuir<sup>1</sup> proposed that  $\text{Na}_{8.0}$  ( $\text{Na}_2\text{O}$  content corrected to 8%  $\text{MgO}$ ) can be used as a proxy for the extent of melting, whereas  $\text{Fe}_{8.0}$  and  $\text{Si}_{8.0}$  indicate the pressure of melting. Their model predicts that basalts from shallow ridges, such as the Reykjanes Ridge or the Red Sea rift, should have low  $\text{Na}_{8.0}$  and  $\text{Si}_{8.0}$  but high  $\text{Fe}_{8.0}$ , whereas deep ridges, such as the Southwest Indian Ridge, should have high  $\text{Na}_{8.0}$  and  $\text{Si}_{8.0}$  and low  $\text{Fe}_{8.0}$ . Substantial scatter in the data is inferred to be the result of major-element heterogeneity in the mantle, high-pressure fractionation, or sampling of liquids produced at different depths without mixing. To filter out such local effects, Klein and Langmuir<sup>4</sup> projected the data onto the global trends and calculated virtual values for iron and silicon ( $\text{Si}_{8.0}^{\text{G}}$  and  $\text{Fe}_{8.0}^{\text{G}}$ ).

The correlations can be interpreted as showing that, as hotter mantle intersects its solidus at greater depth, it produces a taller partially molten column, higher mean pressures and greater extent of melting. The crust above the hot mantle is thicker and because of isostatic compensation, the median valley of the spreading ridge is shallower. This model provides useful insight into the thermal state of the mantle beneath the ridge axis. A

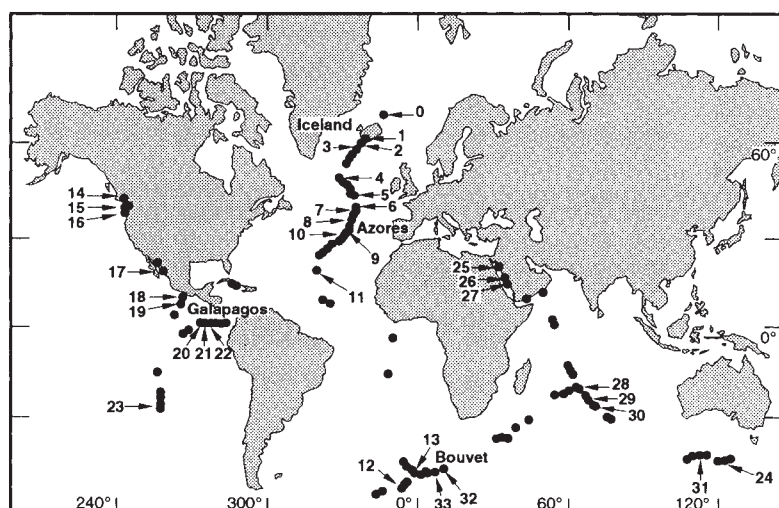


FIG. 1 Location map for data used in this study. Black dots are dredged sites for which major-element chemistry and strontium isotope ratios are available from the literature. Numbers refer to areas where average major-element chemistry and S-wave seismic velocity have been calculated following procedures in refs 1,2,5 (chemistry) and 13 (velocity). Areas 0 to 13 are located in the Atlantic Ocean and the American–Antarctic Ridge<sup>18–27</sup>; 14 to 24 are from the Pacific Ocean<sup>28–36</sup>; 25 to 27 are from the Red Sea<sup>43,44</sup>; 28 to 33 are from the Indian Ocean<sup>37–42</sup>. See Table 1 for coordinates.

temperature range of  $\sim 280$  °C in the subsolidus mantle<sup>2</sup> is required to account for the entire data set presented in ref. 1.

We have compiled published and unpublished major-element analyses for zero-age mid-ocean-ridge basalts (MORB) from the Atlantic, Pacific and Indian oceans and the Red Sea (Fig. 1 and Table 1; refs 18–44, and J. H. Natland, manuscript in preparation). The initial data set contained about 398 samples from 327 stations, but following ref. 1, we used only samples with MgO contents between 5 and 8.5%, which gave us 305 samples from 250 stations. We corrected the compositions for

low-pressure fractionation and averaged the data geographically over segments about 300 km in length. This gave us a total of 34 areas showing most of the chemical variations reported in ref. 1.

During the past 10 years, tomographic investigations on both global and regional scales have revealed significant lateral variations in seismic velocity. These can be attributed to thermal anomalies or chemical heterogeneity. High-quality images of large-scale features derived from long-period seismic data are now available<sup>9–11,16</sup>. In the most recent model, RG5.5<sup>13–15</sup>, the

TABLE 1 Major-element analyses and velocity variations for the selected areas

| Location |         |         | S (N) | Na <sub>8.0</sub> |       | Fe <sub>8.0</sub> <sup>G</sup> |       | Si <sub>8.0</sub> <sup>G</sup> |        | Velocity variation (%) | Location     |         |         | S (N) | Na <sub>8.0</sub> |       | Fe <sub>8.0</sub> <sup>G</sup> |       | Si <sub>8.0</sub> <sup>G</sup> |        | Velocity variation (%) |  |
|----------|---------|---------|-------|-------------------|-------|--------------------------------|-------|--------------------------------|--------|------------------------|--------------|---------|---------|-------|-------------------|-------|--------------------------------|-------|--------------------------------|--------|------------------------|--|
|          |         |         |       | Mean              | s.d.  | Mean                           | s.d.  | Mean                           | s.d.   |                        |              |         |         |       | Mean              | s.d.  | Mean                           | s.d.  | Mean                           | s.d.   |                        |  |
| Atlantic |         |         |       |                   |       |                                |       |                                |        |                        | Pacific      |         |         |       |                   |       |                                |       |                                |        |                        |  |
| 71.33N   | 12.64W  | 0 (6)   | 1.87  | 0.08              | 11.01 | 0.31                           | 50.38 | 0.08                           | 0.111  |                        | 12.44N       | 103.92W | 18 (8)  | 2.91  | 0.31              | 8.19  | 0.74                           | 50.79 | 0.15                           | -0.761 |                        |  |
| 69.15N   | 16.22W  |         |       |                   |       |                                |       |                                |        |                        | 12.87N       | 103.96W |         |       |                   |       |                                |       |                                |        |                        |  |
| 62.59N   | 25.45W  | 1 (10)  | 1.81  | 0.12              | 11.32 | 0.25                           | 50.19 | 0.10                           | -1.670 |                        | 09.91N       | 104.04W | 19 (10) | 2.71  | 0.31              | 8.84  | 0.75                           | 50.75 | 0.15                           | -0.841 |                        |  |
| 63.97N   | 21.74W  |         |       |                   |       |                                |       |                                |        |                        | 10.04N       | 104.73W |         |       |                   |       |                                |       |                                |        |                        |  |
| 61.37N   | 27.40W  | 2 (7)   | 1.96  | 0.09              | 11.03 | 0.18                           | 50.34 | 0.04                           | -1.585 |                        | 01.95N       | 91.18W  | 20 (6)  | 2.25  | 0.20              | 9.96  | 0.51                           | 50.50 | 0.08                           | -0.963 |                        |  |
| 62.37N   | 25.84W  |         |       |                   |       |                                |       |                                |        |                        | 02.60N       | 95.33W  |         |       |                   |       |                                |       |                                |        |                        |  |
| 59.99N   | 29.51W  | 3 (8)   | 1.97  | 0.04              | 10.97 | 0.12                           | 50.33 | 0.04                           | -1.351 |                        | 00.84N       | 89.59W  | 21 (6)  | 2.45  | 0.23              | 9.27  | 0.72                           | 50.50 | 0.10                           | -0.676 |                        |  |
| 61.10N   | 27.88W  |         |       |                   |       |                                |       |                                |        |                        | 01.90N       | 90.95W  |         |       |                   |       |                                |       |                                |        |                        |  |
| 52.66N   | 34.94W  | 4 (6)   | 2.13  | 0.22              | 10.58 | 0.29                           | 50.55 | 0.13                           | -0.641 |                        | 01.04N       | 85.42W  | 22 (10) | 2.10  | 0.21              | 10.18 | 0.62                           | 50.43 | 0.11                           | -0.387 |                        |  |
| 59.99N   | 29.51W  |         |       |                   |       |                                |       |                                |        |                        | 00.84N       | 89.59W  |         |       |                   |       |                                |       |                                |        |                        |  |
| 49.52N   | 28.54W  | 5 (5)   | 2.65  | 0.18              | 8.60  | 0.37                           | 50.83 | 0.14                           | 0.397  |                        | 28.20S       | 112.50W | 23 (7)  | 2.34  | 0.31              | 9.70  | 0.71                           | 50.54 | 0.22                           | -0.965 |                        |  |
| 52.01N   | 29.95W  |         |       |                   |       |                                |       |                                |        |                        | 34.00S       | 104.73W |         |       |                   |       |                                |       |                                |        |                        |  |
| 42.79N   | 29.36W  | 6 (8)   | 2.50  | 0.37              | 9.41  | 0.78                           | 50.72 | 0.51                           | -0.872 |                        | 48.74S       | 127.08E | 24 (6)  | 2.88  | 0.29              | 8.63  | 0.70                           | 50.86 | 0.23                           | -0.574 |                        |  |
| 46.53N   | 27.49W  |         |       |                   |       |                                |       |                                |        |                        | 50.21S       | 137.55E |         |       |                   |       |                                |       |                                |        |                        |  |
| 38.43N   | 30.36W  | 7 (7)   | 2.41  | 0.29              | 8.65  | 0.36                           | 50.61 | 0.16                           | -1.181 |                        | Red Sea      |         |         |       |                   |       |                                |       |                                |        |                        |  |
| 39.63N   | 29.74W  |         |       |                   |       |                                |       |                                |        |                        | 26.26N       | 35.38E  | 25 (4)  | 2.62  | 0.15              | 8.17  | 1.02                           | 50.84 | 0.34                           | -0.446 |                        |  |
| 39.76N   | 29.69W  | 8 (4)   | 2.00  | 0.16              | 10.26 | 0.31                           | 50.50 | 0.11                           | -1.232 |                        | 25.49N       | 36.06E  |         |       |                   |       |                                |       |                                |        |                        |  |
| 40.23N   | 29.59W  |         |       |                   |       |                                |       |                                |        |                        | 21.13N       | 38.08E  | 26 (5)  | 2.18  | 0.18              | 10.55 | 0.61                           | 50.43 | 0.15                           | -1.101 |                        |  |
| 35.84N   | 34.18W  | 9 (6)   | 2.15  | 0.22              | 9.49  | 0.32                           | 50.57 | 0.22                           | -1.003 |                        | 19.07N       | 39.25E  |         |       |                   |       |                                |       |                                |        |                        |  |
| 38.43N   | 30.36W  |         |       |                   |       |                                |       |                                |        |                        | 17.94N       | 40.09E  | 27 (4)  | 1.94  | 0.11              | 10.63 | 0.21                           | 50.55 | 0.07                           | -1.232 |                        |  |
| 35.13N   | 35.00W  | 10 (6)  | 2.20  | 0.18              | 9.26  | 0.18                           | 50.52 | 0.15                           | -0.956 |                        | 18.15N       | 39.99E  |         |       |                   |       |                                |       |                                |        |                        |  |
| 35.33N   | 34.90W  |         |       |                   |       |                                |       |                                |        |                        | Indian Ocean |         |         |       |                   |       |                                |       |                                |        |                        |  |
| 23.52N   | 44.82W  | 11 (10) | 2.64  | 0.24              | 9.17  | 0.45                           | 50.78 | 0.19                           | -0.737 |                        | 24.98S       | 70.01E  | 28 (6)  | 2.98  | 0.12              | 8.02  | 0.32                           | 51.34 | 0.05                           | -0.554 |                        |  |
| 23.62N   | 45.04W  |         |       |                   |       |                                |       |                                |        |                        | 25.78S       | 70.18E  |         |       |                   |       |                                |       |                                |        |                        |  |
| 58.45S   | 15.49W  | 12 (5)  | 3.15  | 0.36              | 7.21  | 0.98                           | 50.95 | 0.56                           | 0.273  |                        | 25.57S       | 70.03E  | 29 (4)  | 2.85  | 0.05              | 8.00  | 0.31                           | 51.18 | 0.06                           | -0.275 |                        |  |
| 59.36S   | 18.06W  |         |       |                   |       |                                |       |                                |        |                        | 26.52S       | 71.92E  |         |       |                   |       |                                |       |                                |        |                        |  |
| 56.28S   | 04.70W  | 13 (5)  | 3.09  | 0.24              | 7.76  | 0.31                           | 51.08 | 0.16                           | 0.047  |                        | 28.90S       | 74.66E  | 30 (4)  | 2.58  | 0.14              | 8.62  | 0.34                           | 50.80 | 0.13                           | 0.220  |                        |  |
| 57.77S   | 07.68W  |         |       |                   |       |                                |       |                                |        |                        | 31.26S       | 76.50E  |         |       |                   |       |                                |       |                                |        |                        |  |
| Pacific  |         |         |       |                   |       |                                |       |                                |        |                        | 49.03S       | 124.00E | 31 (4)  | 3.28  | 0.28              | 7.11  | 0.97                           | 51.35 | 0.24                           | 0.093  |                        |  |
| 47.21N   | 129.12W | 14 (12) | 2.51  | 0.48              | 9.00  | 1.04                           | 50.75 | 0.37                           | -1.013 |                        | 49.81S       | 119.17E |         |       |                   |       |                                |       |                                |        |                        |  |
| 49.05N   | 130.94W |         |       |                   |       |                                |       |                                |        |                        | 54.02S       | 07.26E  | 32 (5)  | 2.93  | 0.20              | 8.90  | 0.38                           | 50.97 | 0.13                           | -0.452 |                        |  |
| 45.95N   | 130.03W | 15 (10) | 2.39  | 0.13              | 9.62  | 0.45                           | 50.59 | 0.07                           | -1.083 |                        | 53.97S       | 07.35E  |         |       |                   |       |                                |       |                                |        |                        |  |
| 47.21N   | 129.12W |         |       |                   |       |                                |       |                                |        |                        | 54.02S       | 03.53E  | 33 (4)  | 2.66  | 0.24              | 8.35  | 0.30                           | 50.88 | 0.03                           | -0.292 |                        |  |
| 44.62N   | 130.39W | 16 (10) | 2.42  | 0.37              | 9.52  | 0.76                           | 50.69 | 0.28                           | -0.592 |                        | 54.42S       | 04.74E  |         |       |                   |       |                                |       |                                |        |                        |  |
| 47.21N   | 129.12W |         |       |                   |       |                                |       |                                |        |                        |              |         |         |       |                   |       |                                |       |                                |        |                        |  |
| 20.84N   | 109.06W | 17 (9)  | 2.59  | 0.27              | 9.12  | 0.68                           | 50.76 | 0.14                           | -0.337 |                        |              |         |         |       |                   |       |                                |       |                                |        |                        |  |
| 20.86N   | 109.08W |         |       |                   |       |                                |       |                                |        |                        |              |         |         |       |                   |       |                                |       |                                |        |                        |  |

Latitude and longitude limits for the selected areas. Na<sub>8.0</sub>, Fe<sub>8.0</sub><sup>G</sup> and Si<sub>8.0</sub><sup>G</sup> correspond to oxides (in wt%) corrected for low-pressure fractionation following the procedure of refs 1, 2 and 5. Velocity variation (%) corresponds to the S-wave velocity deviation (from a reference velocity of 4.4595 km s<sup>–1</sup>) for a harmonic expansion truncated at degree 36 for 110 km depth<sup>13–15</sup>. The error in the velocity variation values is estimated to be about 10% S (N) refer to site number (S) and number of analysis used to calculate the mean values (N). s.d.: standard deviation. See Fig. 1 for reference.



spectrum of S-wave velocity heterogeneity is expanded up to spherical harmonic degree 36 (scale-length  $\sim 1,000$  km). The model is constructed by using Love and Rayleigh waves, in the period range 75–250 s, from 18,000 seismograms associated with 971 events with magnitude larger than 5.5. The results are strongly influenced by mantle velocity variations in the upper 300 km, but are compatible with results for the whole mantle<sup>16,17</sup>. The geographical locations of different velocities are very similar at large wavelength, but their depths vary slightly with the inversion technique by ( $\pm 50$  km in the shallow mantle,  $\pm 100$  km below 300 km). A very common feature of the ridges<sup>12,14</sup> is that velocity patterns are asymmetric below 100 km depth. For each of the 34 regions defined above, we have calculated the S-wave velocity for depths ranging from 66 to 230 km and spherical harmonics truncated at degrees 10, 15 and 36. This enables us to determine whether correlations between tomography and geochemistry are due to short- or long-wavelength anomalies.

We have calculated correlations between the major-element parameters ( $\text{Na}_{8.0}$ ,  $\text{Fe}_{8.0}$ ,  $\text{Si}_{8.0}$ ,  $\text{Fe}_{8.0}^G$  and  $\text{Si}_{8.0}^G$ ) of the basalts erupted at the ridges and the seismic velocity at different depths in the mantle (expressed as the deviation from a reference velocity for each depth). At shallow depths (66–90 km) there is no correlation between the chemical and seismological data. At 100 km depth (and for harmonic expansions truncated at degree 15 or 36) correlation coefficients improve to 0.60,  $-0.61$  and 0.62, respectively, for  $\text{Na}_{8.0}$ ,  $\text{Fe}_{8.0}^G$  and  $\text{Si}_{8.0}^G$ , but remain insignificant for  $\text{Fe}_{8.0}$  and  $\text{Si}_{8.0}$  (Fig. 2 shows the results for truncation at degree 36). These correlation coefficients are significant at better than the 99% confidence level. For a harmonic expansion truncated at degree 10, no correlations were found at any depth from 66 to 230 km; in other words, the correlations result from harmonic degrees greater than 10, which explains why it was not possible to identify such a correlation from previous, lower-resolution tomographic models.

In the bottom panels of Fig. 2, five points (open symbols) are at higher S-wave velocities than would be predicted from the general correlation. Three of these points correspond to sites (4, 5 and 30) located on or within 100 km of a major fracture zone (one large enough to be visible on a 1:10,000,000 map); as fracture zones are poorly resolved by the tomographic model, this may account for the anomalously high velocity. (None of the other sites is so near to a major fracture zone). The resolution of the tomographic model is also poor at high latitudes, where the coverage of seismic ray paths is sparse owing to the small number of large earthquakes and digital seismic stations. This may explain the failure of the data from the Kolbeinsey Ridge (site 0) to fit the correlation.

In contrast to the four sites just discussed, we can see no obvious reason why site 22 should be anomalous. Basalts from the Galapagos area do seem to have anomalous major element compositions compared to normal ridges<sup>2</sup> (anomalously low  $\text{Fe}_{8.0}$  and sometimes  $\text{Na}_{8.0}$ ), but this cannot be the complete explanation, as sites 20 and 21 do fall on the general trend. It is possible that our use of raw values of velocity for each site (instead of averaged values as used in ref. 14) may introduce local effects and errors.

Because we consider these five sites to be anomalous (whatever the reason), we continue the analysis with a reduced data set from which sites 0, 4, 5, 22 and 30 have been removed. Figure 3 shows plots of velocity variation against  $\text{Na}_{8.0}$  (truncated at degree 36) for this new set of 29 regions. At 66 km depth, slow-spreading ridges ( $< 2.5$  cm yr<sup>-1</sup>) are characterized by high S-wave velocity variation, and intermediate to fast ridges by low velocity variation. At greater depths (90 km), slow-spreading ridges show a positive correlation between S-wave velocity variation and  $\text{Na}_{8.0}$ , whereas intermediate- to fast-spreading ridges are shifted towards lower velocities. This produces the apparent scatter in the diagram. At 110 km depth,

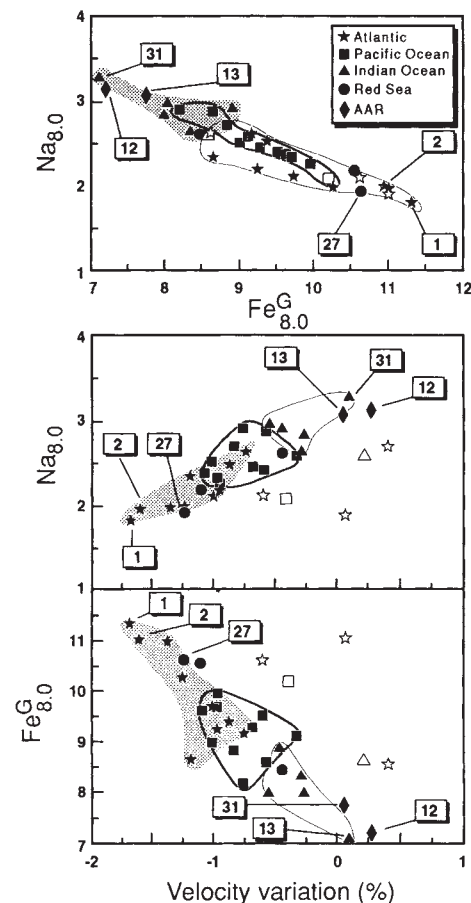


FIG. 2 From top to bottom: plot of  $\text{Na}_{8.0}$  against  $\text{Fe}_{8.0}^G$  and  $\text{Na}_{8.0}$  and  $\text{Fe}_{8.0}^G$  against velocity variation at 110 km depth. Closed symbols correspond to the data used in the calculation; open symbols are sites removed from the calculation (see text for explanation).  $\text{Na}_{8.0}$  and  $\text{Fe}_{8.0}^G$  are sodium and iron content of the rocks corrected for low-pressure fractionation ( $\text{Na}_{8.0}$ ; ref 1) and projected on the global trend to erase local effects ( $\text{Fe}_{8.0}^G$ ; ref 5). Klein and Langmuir suggested that the reverse correlation between  $\text{Na}_{8.0}$  and  $\text{Fe}_{8.0}^G$  is the result of interplay between temperature and pressure within the melting zone. Hot mantle will intersect its solidus at depth, leading to a high  $\text{Fe}_{8.0}^G$  and a low  $\text{Na}_{8.0}$  (areas 1, 2 and 27). In contrast, cold mantle intersects its solidus in the shallower part of the mantle. In this case, basalts erupted onto the sea floor will be characterized by low  $\text{Fe}_{8.0}^G$  and high  $\text{Na}_{8.0}$  content (areas 12, 13 and 31). There is strong correlation between  $\text{Na}_{8.0}$ ,  $\text{Fe}_{8.0}^G$  and seismic velocity variation (here harmonic expansion truncated at degree 36 for 110 km depth). Atlantic (dark grey), Pacific and Indian Ocean fields have been reported. Shallower areas (mostly the Atlantic) have low  $\text{Na}_{8.0}$ , high  $\text{Fe}_{8.0}^G$  and low  $v_s$ . In contrast, the deeper zones (mostly the Indian Ocean) have high  $\text{Na}_{8.0}$ , low  $\text{Fe}_{8.0}^G$  and high  $v_s$ . AAR is the American–Antarctic Ridge.

correlation coefficients increase to  $+0.89$ ,  $-0.89$  and  $+0.84$  respectively for  $\text{Na}_{8.0}$ ,  $\text{Fe}_{8.0}^G$  and  $\text{Si}_{8.0}^G$ . We stress that at these depths there is no correlation between spreading rate and S-wave velocity or  $\text{Na}_{8.0}$ . For depths greater than 170 km, all correlations vanish for the entire range of harmonic expansion used.

Inter-laboratory biases in chemical analyses prevent any definitive conclusion of correlations between silica<sup>3</sup> (and possibly iron) and seismic wave variations. Because the projection scheme onto the global trend essentially forces a correlation of sodium with iron and sodium with silica in the data set as a whole ( $\text{Fe}_{8.0}^G$  and  $\text{Si}_{8.0}^G$ ), it is not surprising that  $\text{Fe}_{8.0}^G$

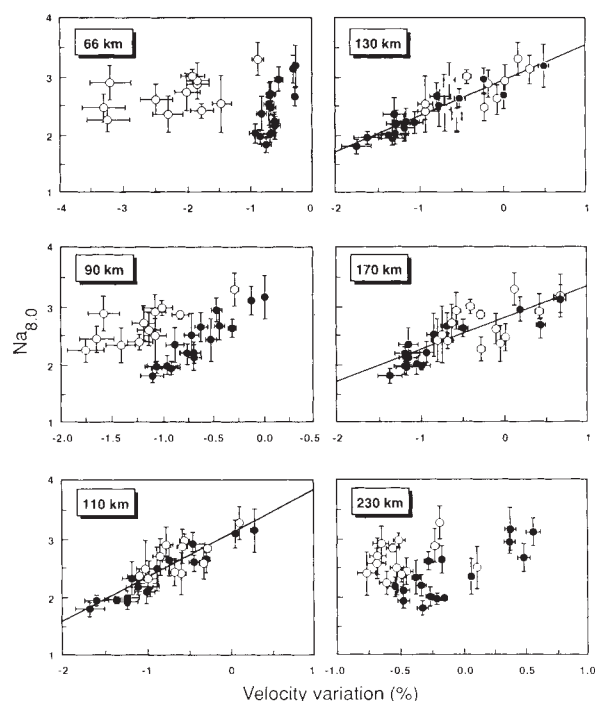


FIG. 3 Average  $\text{Na}_{8.0}$  values ( $\text{Na}_2\text{O}$  wt% corrected for low pressure fractionation) for each of the 29 selected sites plotted against seismic velocity variations (harmonic expansion truncated at degree 36) at different depths within the mantle. There is no global correlation for depths ranging from 66 to 90 km, but the correlation coefficient then increases sharply to +0.89, -0.82 and +0.76, respectively, for 110, 150 and 170 km. Deeper in the mantle, the correlations almost vanish because of the strong asymmetry of the thermal columns beneath most of the ridge axis<sup>12,14</sup>. Shallow ridges such as Reykjanes and the Red Sea have relatively low  $\text{Na}_{8.0}$  and  $\text{Si}_{8.0}^G$ , but  $\text{Fe}_{8.0}^G$  and the ratio  $\text{CaO}/\text{Al}_2\text{O}_3$  are high. These areas (1, 26 and 27) are associated with low seismic velocities. In contrast, deep ridges such as the Australian-Antarctic-Discordance (AAD), and the Southwest Indian Ridge close to the Rodriguez triple junction (area 28) are characterized by higher seismic velocities associated with high  $\text{Na}_{8.0}$  and  $\text{Si}_{8.0}^G$ , and low  $\text{Fe}_{8.0}^G$  and  $\text{CaO}/\text{Al}_2\text{O}_3$ . These chemical variations are the result of thermal variations within the mantle: a range of 280 °C in the subsolidus mantle is required to account for the global chemistry mid-ocean-ridge basalts<sup>2</sup>. On the other hand, lateral heterogeneity in seismic velocities is mostly the result of lateral temperature variation. From these diagrams we calculate that thermal coefficient of shear velocity ( $dV_s/dT$ ) ranges from -0.21 to -0.24  $\text{m s}^{-1} \text{K}^{-1}$ . (See text for explanation.) The reference velocities at 110, 150 and 170 km depth are 4.4595, 4.4447 and 4.4373  $\text{km s}^{-1}$ , respectively. Slow spreading rates ( $<2.5 \text{ cm yr}^{-1}$ ) and intermediate to fast spreading rates (black and white symbols respectively; ref. 45) are indicated. Bars: one standard deviation about the mean  $\text{Na}_{8.0}$  value. Lines: linear regression.

and  $\text{Si}_{8.0}^G$  correlate with seismic velocity variation even if  $\text{Fe}_{8.0}$  and  $\text{Si}_{8.0}$  do not.

Major-element chemistry corrected for low-pressure fractionation and S-wave velocity variations are strongly related for depths ranging from 110 to 170 km. These correlations are apparent only for spherical harmonics truncated at degrees 15 and 36 (scale-lengths of 2,400 and 1,000 km).

As the spreading rate increases ( $> 2.5 \text{ cm yr}^{-1}$ ), a greater fraction of melt is retained at shallower depths in the mantle<sup>14</sup>. This produces slower S-wave velocities and increases the scatter of the data (see Fig. 3 at 90 km). This explains why correlations are insignificant at 66 km depth and hold only for slow-spreading ridges at 90 km. Between 100 and 170 km depth, it is unlikely that melt occurs within the mantle, so most of the velocity variation must be due to thermal variation or chemical heterogeneity<sup>46</sup>. From the slopes in Fig. 3 for the 29 regions selected, we calculated the thermal coefficient of shear velocity

( $dV_s/dT$ ) by converting  $\text{Na}_{8.0}$  variations to temperature differences (see ref. 2 for calculation procedure). The results are about -0.21  $\text{m s}^{-1} \text{K}^{-1}$  and -0.24  $\text{m s}^{-1} \text{K}^{-1}$  between 110 and 170 km depth. These values are compatible with the gradients inferred from laboratory experiments<sup>47</sup> on single crystals.

The main cause of the correlations thus seems to be thermal variations in a liquid-free mantle. Seismic velocity variations are more influenced by temperature than composition at these depths. This conclusion is consistent with those of Bass and Anderson<sup>48</sup> and Yan *et al.*<sup>45</sup>. Below 170 km, the thermal structures are strongly asymmetric, so the basalt chemistry and seismic S-wave velocities are no longer correlated.

Overall, seismic velocity variations and major-element compositions of basalt give coherent results which strongly point to a common thermal origin. To test this further, our approach should be extended to other geodynamic environments. □

Received 1 February; accepted 26 May 1993.

- Klein, E. M. & Langmuir, C. H. *J. geophys. Res.* **92**, 8089–8115 (1987).
- Langmuir, C. H., Klein, E. M. & Plank T. *J. geophys. Res.* (in the press).
- McKenzie, D. & Bickle, M. J. *J. Petrol.* **29**, 625–679 (1988).
- Brodholt J. P. & Batiza, R. *J. geophys. Res.* **94**, 4231–4239 (1989).
- Klein, E. M. & Langmuir, C. H. *J. geophys. Res.* **94**, 4241–4252 (1989).
- Masters, G., Jordan, T. H., Silver, P. G. & Gilbert, F. *Nature* **298**, 609–613 (1982).
- Nakanishi, I. & Anderson, D. L. *Bull. seism. Soc. Am.* **72**, 1185–1194 (1982); *Geophys. J. R. astr. Soc.* **78**, 573–617 (1984).
- Nataf, H. C., Nakanishi, I. & Anderson D. L. *Geophys. Res. Lett.* **11**, 109–112 (1984); *J. geophys. Res.* **91**, 7261–7307 (1986).
- Woodhouse, J. H. & Dziewonski A. M. *J. geophys. Res.* **89**, 5953–5986 (1984).
- Tanimoto, T. & Anderson, D. L. *J. geophys. Res.* **90**, 1842–1858 (1985).
- Tanimoto, T. *J. Geophys. J. R. astr. Soc.* **82**, 105–123 (1986); **84**, 49–69 (1986).
- Montagner, J. P. & Tanimoto, T. *J. geophys. Res.* **95**, 4797–4819 (1990).
- Zhang, Y. S. & Tanimoto, T. *J. Phys. Earth planet Int.* **66**, 160–202 (1991).
- Zhang, Y. S. & Tanimoto, T. *Nature* **355**, 45–49 (1992).
- Zhang, Y. S. & Tanimoto, T. *J. J. geophys. Res.* (submitted).
- Tanimoto, T. *J. Geophys. J. int.* **100**, 327–336 (1990).
- Su, W. J. & Dziewonski, A. M. *Nature* **352**, 121–126 (1992).
- Engel, A. E. J., Engel, C. G. & Havens, R. G. *Geol. Soc. Am. Bull.* **76**, 719–734 (1965).
- Langmuir, C. H., Bender, J. F., Bence, A. E. & Hanson, G. N. *Earth planet Sci. Lett.* **36**, 133–156 (1977).
- Sigurdsson, H. *J. geophys. Res.* **86**, 9483–9502 (1981).
- Bryan, W. B. *et al. J. geophys. Res.* **86**, 11815–11836 (1976).
- Bryan, W. B. & Moore, J. G. *Geol. Soc. Bull.* **88**, 556–570 (1977).
- Cohen, R. S. & O'Nions R. K. *J. Petrol.* **23**, 299–324 (1982).
- Jackobsson, S. P., Jonsson, J. & Shido, F. *J. Petrol.* **19**, 669–705 (1978).
- Humphris, S. E., Thompson G., Schilling, J. G. & Kinsley, R. H. *Geochim. cosmochim. Acta* **49**, 1445–1464 (1985).
- Le Roex, A. P., Dick, H. J. B., Gulen, L., Reid, A. M. & Erlank, A. J. *Geochim. cosmochim. Acta* **51**, 541–555 (1987).

- Schilling J. G., Zajac, M., Evans, R., Johnson T., White, W., Devine, J. D., Kinsley, R. *Am. J. Sci.* **283**, 510–586 (1983).
- Schilling, J. G., Kingsley, R. H. & Devine J. D. *Nature* **242**, 565–571 (1972).
- McDougall, J. D. & Lugmair, G. W. *Nature* **313**, 209–211 (1985).
- Printzof, A., Lewin, E. & Allegre C. J. *Earth planet. Sci. Lett.* **92**, 189–206 (1989).
- Fornari D. J. *et al. Earth planet Sci. Lett.* **89**, 63–83 (1988).
- Cohen, R. S., Evensen, N. M., Hamilton, P. J. & O'Nions, R. K. *Nature* **283**, 149–153 (1980).
- Kay, R., Hubbard, N. J. & Gast, P. W. *J. geophys. Res.* **75**, 1585–1595 (1972).
- Puchelt, H. & Emmermann, R. *Chem. Geol.* **38**, 39–56 (1983).
- Eaby, J., Clague, D. A. & Delaney, J. R. *J. geophys. Res.* **89**, 7883–7890 (1984).
- Klein, E. M., Langmuir, C. H. & Staudigel, H. J. *geophys. Res.* **96**, 2089–2108 (1991).
- Dosso, L., Bougault, H., Beuzart, P., Calvez, J. Y. & Joron J. L. *Earth Planet. Sci. Lett.* **88**, 47–59 (1988).
- Price, R. C., Kennedy, A. K., Riggs-Sneeringer, M. & Frey, F. A. *Earth planet. Sci. Lett.* **78**, 379–396 (1986).
- Le Roex, A. P. *et al. J. Petrol.* **24**, 267–318 (1983).
- Le Roex, A. P. *et al. Contrib. Miner. Petrol.* **90**, 367–380 (1985).
- Humler, E. & Whitechurch, H. *Earth planet. Sci. Lett.* **88**, 169–181 (1988).
- Humler, E. thesis, Univ. Strasbourg (1989).
- Altherr, R., Henjes-Kunst, F. & Puchelt, H. *J. geophys. Res.* **92**, 4881–4893 (1987).
- Eissen, J. P. *et al. J. Petrol.* **30**, 791–839 (1989).
- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. *Geophys. J. int.* **101**, 425–478 (1990).
- Yan, B., Graham, E. K. & Furlong, K. P. *Geophys. Res. Lett.* **16**, 449–452 (1989).
- Anderson, O. L., Isaak, D. & Oda, H. *Rev. Geophys.* **30**, 57–90 (1992).
- Bass, J. D. & Anderson, D. L. *Geophys. Res. Lett.* **11**, 237–240 (1984).

ACKNOWLEDGEMENTS. This paper benefited from discussions with T. Tanimoto, B. Dupré, Ph. Sarda, E. Lewin, C. J. Allegre, A. W. Hofmann and G. Wagerberg, and critical comments from C. H. Langmuir and E. Klein. We thank J. Dyon for drafting. This work was supported in part by the NSF at Lamont Doherty.



# Chaos reduces species extinction by amplifying local population noise

J. C. Allen\*, W. M. Schaffer† & D. Rosko†

\* Department of Entomology and Nematology, University of Florida, Gainesville, Florida 32611-0740, USA

† Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA

IN the mid-1970s, theoretical ecologists were responsible for stimulating interest in nonlinear dynamics and chaos<sup>1-3</sup>. Ironically, the importance of chaos in ecology itself remains controversial<sup>4-17</sup>. Proponents of ecological chaos point to its ubiquity in mathematical models and to various empirical findings<sup>15,16,18</sup>. Sceptics<sup>12,19,20</sup> maintain that the models are unrealistic and that the experimental evidence is equally consistent with stochastic models. More generally, it has been argued<sup>9,11,21-23</sup> that interdemographic selection and/or enhanced rates of species extinction will eliminate populations and species that evolve into chaotic regions of parameter space. Fundamental to this opinion is the belief<sup>24,25</sup> that violent oscillations and low minimum population densities are inevitable correlates of the chaotic state. In fact, rarity is not a necessary consequence of complex dynamical behaviour<sup>26,27</sup>. But even when chaos is associated with frequent rarity, its consequences to survival are necessarily deleterious only in the case of species composed of a single population. Of course, the majority of real world species (for example, most insects) consist of multiple populations weakly coupled by migration, and in this circumstance chaos can actually reduce the probability of extinction. Here we show that although low densities lead to more frequent extinction at the local level<sup>28</sup>, the decorrelating effect of chaotic oscillations reduces the degree of synchrony among populations and thus the likelihood that all are simultaneously extinguished.

Consider a group of spatially distributed, interbreeding populations, that is, a metapopulation<sup>29-31</sup>, linked by migration and subject to external perturbations. In the absence of migration and disturbance, we suppose that each population renews itself according to a one-dimensional difference equation

$$x_{i,t+1} = f(x_{i,t}) \quad (1)$$

Here,  $x$  denotes density and the subscripts  $i$  and  $t$  index population and time.

To study the effects of population dynamics on species extinction rates, we modify equation (1) to include migration and environmental disturbance. To incorporate migration, we write

$$x_{i,t+1} = f[(1-m)x_{i,t} + m\bar{x}_t] \quad (2)$$

where  $m$  is the emigration rate,  $m x_{i,t}$  is the number of individuals dispersing from the  $i$ th population, and  $\bar{x}_t$  is the mean density taken over all populations at time  $t$ . For notational convenience, we define the migration operator,  $x'_{i,t} = (1-m)x_{i,t} + \bar{x}_t$ .

To incorporate environmental perturbations, we distinguish two kinds of noise: (1) Global noise. To model pervasive disturbances, such as variations in regional weather patterns, we introduce the concept of global noise,  $z_t$ . Typically, we assume that global perturbations occur infrequently, that is, that the probability,  $P$ , of a global event is small, but with large effect. We further presume that a given perturbation affects all populations equally. Global noise is consequently a correlating influence tending to synchronize local population dynamics. (2) Local noise. To model disturbances affecting individual populations, we introduce the concept of local noise,  $u_{i,t}$ . We assume that local disturbances occur every generation and that the perturbations affecting different populations are independent. Local noise is thus a decorrelating force which increases variation among populations.

In the calculations that follow, we assume that  $z_t$  is either normally distributed with mean 0 and standard deviation,  $s_G$  (abbreviated  $N(0, s_G)$ ) or uniformly distributed on  $[0, a_G]$  (abbreviated  $U(0, a_G)$ ). We further assume that the local perturbations,  $u_{i,t}$ , are distributed as  $U(-a_L, a_L)$ .

We first introduce noise as simple additive random perturbations to the one-dimensional map as

$$x_{i,t+1} = f(x'_{i,t}) + u_{i,t} + z_t \quad (3)$$

In implementing this equation, we assume that local noise is distributed as  $U(-a_L, a_L)$ , and global noise is distributed as  $N(0, s_G)$  and is applied with probability  $P$  and equal to zero with probability  $(1-P)$ .

Alternatively, we consider

$$x_{i,t+1} = f(x'_{i,t} + u_{i,t})z_t \quad (4)$$

where the local noise enters through migration as an additive  $U(-a_L, a_L)$  variable and global noise is a random mortality factor distributed as  $U(0, a_G)$ , ( $a_G \leq 1$ ).

Finally, we let the local noise enter through the reproductive rate as

$$x_{i,t+1} = f(x'_{i,t}, r_{i,t})z_t \quad (5)$$

where  $r_{i,t}$  is distributed as  $U(r-a_L, r+a_L)$  and  $z_t$  is the same as in equation (4). In equations (4) and (5), global noise is applied with probability  $P$  and equal to 1 with probability  $(1-P)$ .

To what extent do population dynamics affect the balance between correlating and decorrelating perturbations from without? Recall that chaotic systems exhibit 'sensitivity to initial conditions'<sup>32</sup>, that is, nearby trajectories diverge exponentially, whereas systems exhibiting steady state or periodic behaviour have no such tendency. This leads us to conjecture that, in the absence of other complicating factors, species extinction rates should decline as one goes from regular to chaotic motion, that is, as the parameters in equation (1) are 'tuned' in an appropriate fashion. To test this prediction, we implemented equations (3), (4) and (5) as follows. (1) For the function  $f(x_{i,t})$ , we used the Ricker<sup>33</sup> equation

$$x_{i,t+1} = x_{i,t} \exp(r(1-x_{i,t})) \quad (6)$$

and the logistic map<sup>34</sup>,

$$x_{i,t+1} = r x_{i,t} (1-x_{i,t}) \quad (7)$$

Here  $r$  is the maximum per generation rate at which the population multiplies. In both cases, increasing  $r$  induces period-doubling bifurcations to chaos<sup>1-3,34</sup> and a concomitant decline in minimum population densities. Equations (6) and (7) thus conform to the rarity assumption noted above. (2) The parameter  $r$  was varied from 1.75 to 3.5 for the Ricker map and 2.95 to 4.0 for the logistic. Typically, we computed extinction rates for each pixel on the computer screen over the range of  $r$  (556 parameter values). (3) Global noise was introduced as follows: in equation (3),  $z_t$  was  $N(0, 1)$  for the Ricker map, and  $N(0, 0.25)$  for the logistic; in equation (4),  $z_t$  was  $U(0, 0.1)$  for both maps; in equation (5)  $z_t$  was  $U(0, 0.5)$ . Local noise was taken to be small ( $a_L < 0.10$ ). (4) Extinctions were recorded as follows: in equation (3), a population extinction was noted whenever the density of a single population dropped to  $x_i \leq 0$ . In equations (4) and (5),  $x_i < 0.01$  and  $x_i < 0.001$ , respectively, were interpreted as an extinction. A species extinction was recorded whenever all  $N$  populations became extinct. (5) Following local extinctions, populations were reset to zero and recolonized by immigration. Following a species extinction, all populations were re-initialized at

$$x_{i,0} = 0.025 + u_{i,0} \quad (8)$$

Figures 1 and 2 summarize our findings for equation (3) for the Ricker and logistic maps with 10 populations and weak migration ( $m=0.01$ ), with 50,000 generations per value of  $r$ .

Here we superimpose rates of population and species extinction on bifurcation diagrams for the respective map. This allows us to compare extinction in the metapopulation with single population dynamics. We display results for two probabilities of global perturbation ( $P=0.05$  and  $0.10$ ), and four levels of local noise. We call attention to the following: (1) Generally speaking, local extinction rates increase as one moves from stable to chaotic dynamics. Thus at the local level, extinction probabilities vary in the manner predicted by opponents of ecological chaos. (2) For the case of  $a_L=0$ , that is, no local noise, population and species extinction rates are identical. This is because there is no variation among populations upon which sensitivity to initial conditions can act, so  $u_{i,t}=0$  for all  $i$  and  $t$ , and we are effectively dealing with a single population. (3) In the presence of local noise, species extinction rates decline in the chaotic region. For a small value of  $a_L$ , species extinction rates go back up in the periodic windows. This suggests that it really is the chaos, namely the divergence of nearby itineraries, as opposed to the overall shape of  $f(\cdot)$ , that determines the system's behaviour. (4) With increasing amounts of local noise, the protection afforded by chaos is enhanced. At the same time, the effect of the periodic windows disappears. This reflects the fact<sup>35</sup> that motion in the windows now corresponds to noise-stabilized chaotic transients. (5) Amplification of local noise by chaos is not very sensitive to

the method by which the noise is introduced. When noise is introduced according to equations (4) and (5), we find a reduction of species extinction rates similar to equation (3) (Fig. 3). We believe the effect to be fairly robust. For moderate amounts of local and global noise, chaos will amplify the local noise and increase the heterogeneity of population densities.

With noise modelled according to equation (3), we have obtained the following additional results, to be reported in detail elsewhere (D.R., W.M.S. and J.C.A., manuscript in preparation). (1) The protective effect of chaos depends on weak coupling by migration ( $m<0.1$ ). Zero migration results in rapid species extinction as local extinctions cannot be recolonized. Conversely, with high levels of migration, we are back to the case of a single population. (2) We assume that individuals dispersing from a particular population are equally likely to wind up in any other population. Replacing this assumption with schemes that incorporate explicit spatial arrangements, for example, diffusive spread in a one-dimensional environment, has essentially no effect on the pattern shown in Fig. 1. Thus, although the overall probability of migration has a significant effect on extinction, the particular dispersal pattern is relatively inconsequential. (3) In the case of the logistic map, local populations go extinct with probability one, for  $r>4$ . In our system, these populations can be recolonized by migration, and  $r$  values greater than 4 do

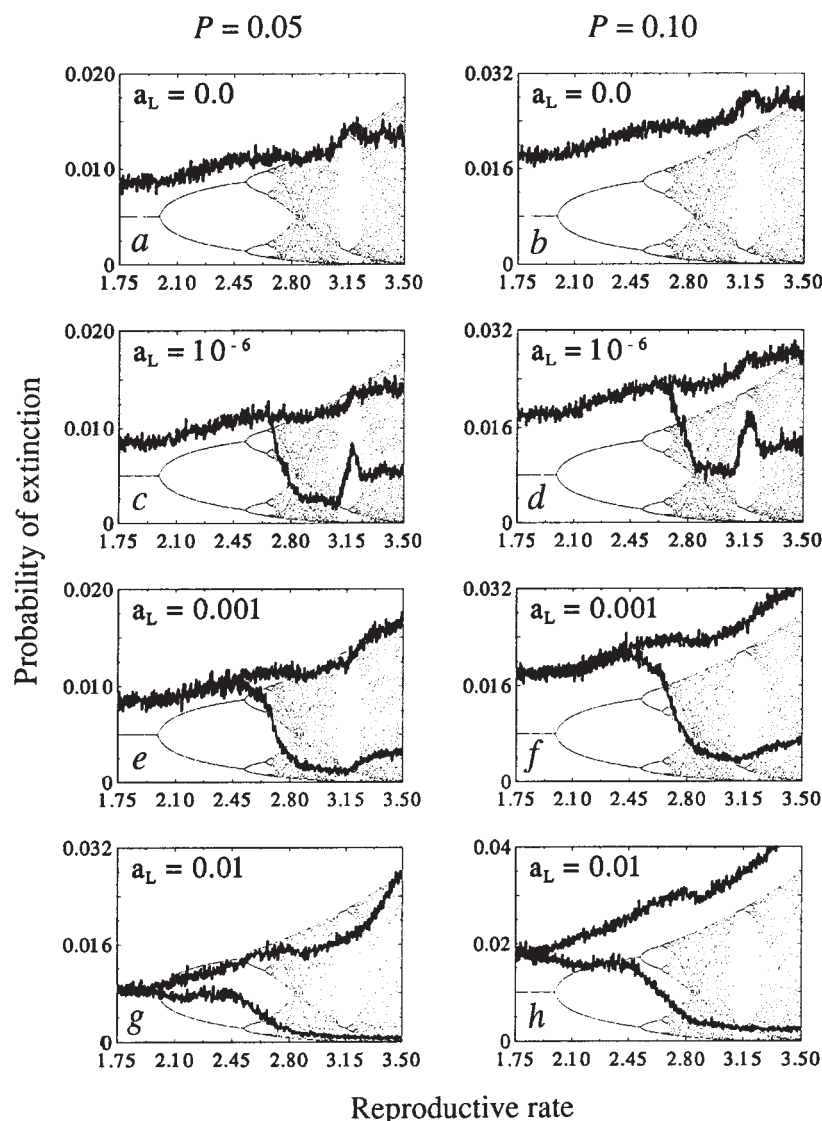


FIG. 1 Probabilities of local and species extinction for 10 Ricker populations in which noise is introduced by equation (3). Extinction rates are superimposed on bifurcation diagrams for a single population in the absence of noise and migration. Where the lines diverge (c-h), the upper line is the local extinction rate, and the lower line, the species extinction rate. Fifty-thousand generations were computed for each of 556  $r$  values on  $[1.75, 3.5]$ . The migration rate,  $m$ , was set to 0.01, and the frequency of global perturbations to  $p=0.05$  (left) and  $0.10$  (right). Global noise was distributed as  $N(0, 1)$ . Local noise was distributed as  $U(-a_L, a_L)$ . a, b, In the case of no local noise ( $a_L=0.0$ ), the probability of species and local extinction are identical. c, d, For small amounts of local noise ( $a_L=10^{-6}$ ), there is a sharp drop in species extinction rates in the chaotic region followed by an increase in the period-three window. e-h, For greater local noise ( $a_L=0.001$  and  $0.01$ ) the effects of the period-three window are 'fuzzed out' by noise-stabilized chaotic transients.



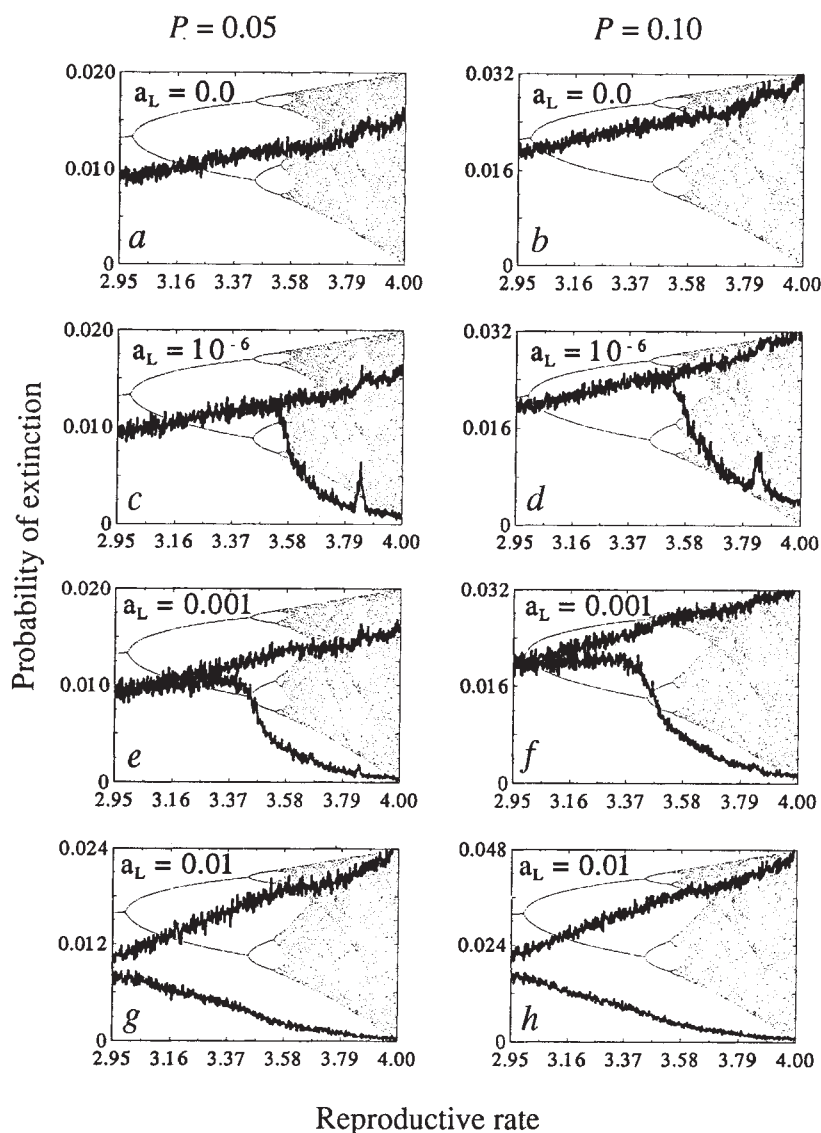


FIG. 2 As Fig. 1, but using the logistic map, equation (7), for  $f(\cdot)$ . Global noise was distributed as  $N(0, 0.25)$ .

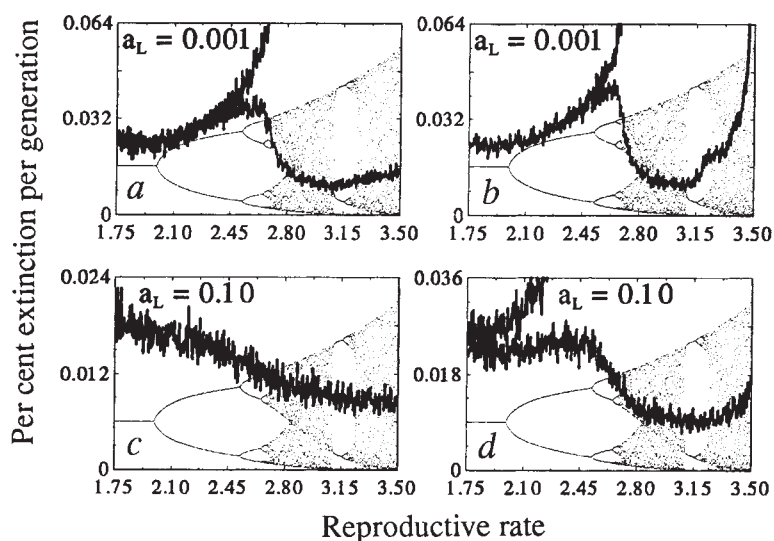


FIG. 3 Extinction rates for equations (4) and (5) where local noise enters through migration (a, c) and reproduction (b, d), respectively. Global noise is a random survival factor distributed as  $U(0, 0.5)$  applied with probability 0.10, and migration rate ( $m$ ) is 0.01. Two levels of local noise ( $a_L = 0.001, 0.10$ ) are shown. The upper line going off scale is the local extinction rate (this line is completely off-scale in c). The lower line is the global or species extinction rate. Only in the case of noise entering through reproductive rate at the lowest level of local noise (b) and highest reproductive rate does the protection afforded by chaos begin to fail. When chaos has higher levels of local noise to amplify (d), the protection is maintained.

not necessarily cause species extinction. For suitable parameter values we have observed reductions in extinction rates up to  $r = 12.0$ . (4) Log-log plots of mean time to extinction,  $\bar{T}$ , versus mean time between perturbations,  $\bar{\tau} = 1/p$ , suggest scaling laws of the form

$$\bar{T} = A\bar{\tau}^B \quad (9)$$

For small  $\bar{\tau}$ , the exponent,  $B$ , is determined by the rate of trajectorial separation, that is, by the magnitude of the positive Lyapunov exponent. For large  $\bar{\tau}$ ,  $B$  approaches an expected value of unity as the system relaxes to its invariant measure<sup>36</sup>. In this case, the chaotic systems have extinction times ( $\bar{T}$ ) 100 to 1,000 times greater than their non-chaotic counterparts.

The foregoing calculations suggest that species extinction is not the inevitable consequence of chaotic population dynamics that some would maintain. On the contrary, chaos can provide significant protection against extinction by increasing the degree of asynchrony among local populations. This result is consistent with recent conclusions<sup>37</sup> that subdividing a population can stabilize systems that would otherwise go extinct with probability one. In those experiments, extinction resulted from deterministic forces, whereas here, the responsible factors are stochastic. The essential mechanism is nonetheless the same: local populations become decorrelated so that when one goes extinct, recolonization by migrants prevents extinction of the whole. We have extended this discussion by pointing out that the stabilizing effects of subdivision and migration are enhanced by chaos. One possible example of 'chaos-mediated' survival is the persistence of highly contagious diseases, such as measles, in human populations below what is called the "critical community size"<sup>38</sup>. Here, as in the ecological case considered above, unstable (chaotic?) dynamics induce asynchrony among local populations, thereby reducing the pathogen's chance of extinction. Potentially interesting extensions of our results include (1) the search for circumstances in which chaos reduces extinction rates at both the local and species levels, and (2) examination of the effect of chaos on extinction in spatially distributed systems of interacting species. □

36. Lichtenberg, A. J. & Leiberman, M. A. *Regular and Stochastic Motion* (Springer, New York, 1983).
37. Yorke, J. A., Nathanson, N., Pianigiani, G. & Martin, J. *Am. J. Epidem.* **109**, 103–122 (1979).
38. Hassell, M. P., Comins, H. N. & May, R. M. *Nature* **353**, 255–258 (1991).

ACKNOWLEDGEMENTS. We thank P. Chesson, G. Fox, J. H. Frank, B. Kendall, J. Keesling, M. Kot, J. Logan, R. May, S. Schwinning, G. Sugihara, C. Tidd, P. Turchin, T. J. Walker and A. Winfree for comments and discussion. This work was supported by grants from USDA to J.C.A. and NIH to W.M.S.

## Population oscillations of boreal rodents: regulation by mustelid predators leads to chaos

Ilkka Hanski\*, Peter Turchin†, Erkki Korpimäki‡ & Heikki Henttonen§

\* Department of Zoology, Division of Ecology, PO Box 17 (P. Rautatiekatu 13), SF-00014 University of Helsinki, Finland  
† Southern Forest Experiment Station, 2500 Shreveport Highway, Pineville, Louisiana 71360, USA

‡ Laboratory of Ecological Zoology, Department of Biology, University of Turku, SF-20500 Turku, Finland

§ Department of Forest Ecology, Finnish Forest Research Unit, PO Box 18, SF-01301 Vantaa, Finland

THE four-year cycle of microtine rodents in boreal and arctic regions was first described in 1924 (ref. 1). Competing hypotheses on the mechanisms underlying the small mammal cycle have been extensively tested<sup>2–5</sup>, but so far the sustained rodent oscillations are unexplained. Here we use two mutually supportive approaches to investigate this question. First, building on studies of the interaction between rodents and their mustelid predators<sup>6–9</sup>, we construct a predator–prey model with seasonality. Second, we use a new technique of nonlinear analysis<sup>10,11</sup> to examine empirical time-series data, and compare them with the model dynamics. The model parameterized with field data predicts dynamics that closely resemble the observed dynamics of boreal rodent populations. Both the predicted and observed dynamics are chaotic, albeit with a statistically significant periodic component. Our results suggest that the multiannual oscillations of rodent populations in Fennoscandia are due to delayed density dependence imposed by mustelid predators, and are chaotic.

As a starting point we use the general predator–prey model<sup>12</sup>

$$\frac{dN}{dt} = rN \left( 1 - \frac{N}{K} \right) - \frac{cPN}{N+D} \quad (1)$$

$$\frac{dP}{dt} = vP \left( 1 - \frac{qP}{N} \right) \quad (2)$$

where  $N$  and  $P$  are the prey and predator densities (numbers of individuals per hectare). Numerical responses of both species are logistic, with  $r$  and  $v$  being the intrinsic growth rates, and  $K$  and  $N/q$  the 'carrying capacities', respectively. The predator–prey interaction is modelled by a type-2 functional response with parameters  $c$  and  $D$ .

Seasonality exerts a strong influence on the dynamics of mammals in boreal and arctic regions. We include seasonality in the model as follows. In Fennoscandia, the potential summer breeding season for rodents is about 6 months long<sup>13</sup>, hence we assume equally long summer and winter seasons. During summer, the rodent population breeds continuously, so we use equation (1). Mustelids breed only when prey density exceeds a critical threshold<sup>14–16</sup>, say  $N_{crit}$ . In the model, predator numbers change

Received 30 November 1992; accepted 26 April 1993.

1. May, R. M. *Science* **186**, 645–647 (1974).
2. May, R. M. *Nature* **216**, 459–467 (1976).
3. May, R. M. & Oster, G. F. *Am. Nat.* **110**, 573–599 (1976).
4. Hassell, M. P., Lawton, J. H. & May, R. M. *J. Anim. Ecol.* **45**, 473–486 (1976).
5. Schaffer, W. M. & Kot, M. *Trends Ecol. Evol.* **1**, 58–63 (1986).
6. Nisbet, R. M., Blythe, S. & Gurney, W. S. C. *Trends Ecol. Evol.* **4**, 238–239 (1989).
7. Mani, G. S. *Trends Ecol. Evol.* **4**, 239–240 (1989).
8. Lomnicki, A. *Trends Ecol. Evol.* **4**, 239 (1989).
9. Beryman, A. A. in *Chaos in Ecology* (eds Logan, J. A. & Hain, F. P.) 23–38 (VA Exp. Sta. Inf. Series 91–93, Blacksburg, VA, 1991).
10. Morris, W. F. *Ecology* **71**, 1849–1862 (1990).
11. Beryman, A. A. & Millstein, J. A. *Trends Ecol. Evol.* **4**, 26–28 (1989).
12. Ellner, S. in *Chaos in Ecology* (eds Logan, J. A. & Hain, F. P.) 63–90 (VA Exp. Sta. Inf. Series 91–93, Blacksburg, VA, 1991).
13. Sugihara, G. & May, R. M. *Nature* **344**, 734–741 (1990).
14. Turchin, P. *Nature* **344**, 660–663 (1990).
15. Turchin, P. in *Chaos in Ecology* (eds Logan, J. A. & Hain, F. P.) 39–62 (VA Exp. Sta. Inf. Series 91–93, Blacksburg, VA, 1991).
16. Turchin, P. & Taylor, A. *Ecology* **73**, 289–305 (1992).
17. Logan, J. A. & Allen, J. C. A. *Rev. Ent.* **37**, 455–477 (1992).
18. Tillman, D. & Wedin, D. *Nature* **353**, 653–655 (1991).
19. Casdagli, M. *Physica D* **35**, 335–356 (1989).
20. Ruelle, D. *Proc. R. Soc. Lond. A* **427**, 241–248 (1990).
21. Thomas, W. R., Pomerantz, M. J. & Gilpin, M. E. *Ecology* **61**, 13–17 (1980).
22. Mueller, L. D. & Ayala, F. J. *Ecology* **62**, 1148–1154 (1981).
23. Philippi, T. E., Carpenter, M. P., Case, T. J. & Gilpin, M. E. *Ecology* **68**, 154–159 (1987).
24. Vandermeer, J. *Ecology* **63**, 1167–1168 (1982).
25. Vandermeer, J. *Theor. Pop. Biol.* **22**, 17–27 (1982).
26. Rogers, T. D. *Math. Biosci.* **72**, 13–17 (1984).
27. Milton, J. G. & Belair, J. *Theor. Pop. Biol.* **37**, 273–289 (1990).
28. Pimm, S. L., Jones, J. L. & Diamond, J. *Am. Nat.* **132**, 757–785 (1988).
29. Iwasa, Y. & Roughgarden, J. *Theor. Pop. Biol.* **30**, 194–214 (1986).
30. Roughgarden, J. & Iwasa, Y. *Theor. Pop. Biol.* **29**, 235–261 (1986).
31. Gilpin, M. E. & Hanski, I. *Metapopulation Dynamics: Empirical and Theoretical Investigations* (Academic, London, 1991).
32. Ruelle, D. *Ann. N. Y. Acad. Sci.* **316**, 408–416 (1979).
33. Ricker, W. E. *J. Fish. Res. Bd. Canada* **11**, 559–623 (1954).
34. Collet, P. & Eckmann, J.-P. *Iterated Maps on the Interval as Dynamical Systems* (Birkhäuser, Boston, 1980).
35. Crutchfield, J. P., Farmer, J. D. & Huberman, B. A. *Phys. Rev.* **92**, 45–82 (1982).



according to equation (2) if  $N > N_{\text{crit}}$ , otherwise the predator population declines exponentially, according to

$$\frac{dP}{dt} = d_{\text{high}}P \quad (3)$$

With these assumptions, predator growth rate is a function of the prey/predator ratio when densities are high, thus taking into account competitive interactions among the predators<sup>17</sup>, but at low densities, predators decline, avoiding a problem in the standard ratio-dependent theory<sup>18</sup>.

During winter, food availability for rodents is lower than in summer, as there is no plant growth, the quality of the forage deteriorates, and some foraging sites become inaccessible under snow. We use equation (1) for the rodent population in winter, but with  $r$  and  $K$  replaced by (the lower values of)  $r'$  and  $K'$ . These assumptions lead to population growth in winters when the autumn density is low, but to a population decrease in other winters. This is what happens in *Microtus* voles, the key prey species for mustelids<sup>8,19,20</sup>; *Microtus* typically breed in the winter following the low phase of the rodent cycle<sup>21,22</sup>. Winter breeding has never been observed in mustelids in Fennoscandia, hence we assume none in the model. But rate of mortality in the predator population is evidently related to prey density. We assume an exponential decline in the predator population during winter, with the rate constant depending on prey density. If  $N > N_{\text{crit}}$ , the rate constant has a value  $d_{\text{low}}$ , otherwise the same (higher) value  $d_{\text{high}}$  as in summers when rodents are scarce.

Equations (1) and (2) describe a two-dimensional continuous-time model, which can exhibit only two kinds of qualitative behaviour: a stable equilibrium point or a limit cycle<sup>7,12</sup>. In the modified version with seasonality, more complex dynamics are possible<sup>23,24</sup>, and it is easy to pick up parameter values to generate stable dynamics, limit cycles and chaos. The critical question is what kind of dynamics are predicted by parameter values estimated from real data. If the model with field-estimated parameters generates population trajectories that are quantitatively similar to the observed dynamics, then the predation hypothesis explaining microtine oscillations is supported; otherwise it is rejected.

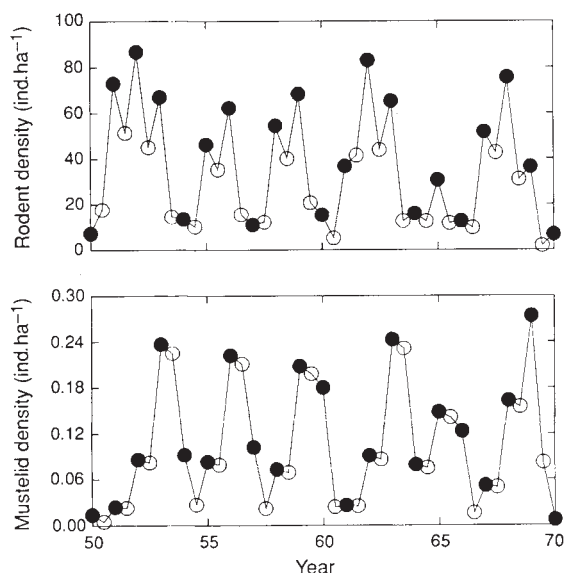


FIG. 1 An example of predicted rodent-mustelid dynamics, generated with the median values of the three parameters that were varied ( $q=100$ ,  $D=5$  and  $K'=50$ ; for explanation see Table 1 legend). Upper panel shows changes in rodent density; lower panel, in mustelid density. Open symbols give spring densities, filled symbols autumn densities. The values on the x axis are generations since the beginning of the simulation.

We have estimated the parameter values based primarily on field data collected by E.K. and coworkers in western Finland in 1977–1992 (Table 1). Three parameters are hard to estimate, and we have numerically explored the behaviour of the model for three values of each of the three parameters, a total of 27

TABLE 1 Analysis of model output

| $q$ | $D$  | $K'$ | $s$  | $T$ | ACF[T] | LE1   | Dynamics     |
|-----|------|------|------|-----|--------|-------|--------------|
| 50  | 5.0  | 25   | 0.33 | 4   | 0.30   | 0.60  | Chaos        |
| 50  | 5.0  | 50   | 0.62 | 8   | 0.38   | 0.59  | Chaos        |
| 50  | 5.0  | 75   | 0.84 | 10  | 0.32   | 0.39  | Chaos*       |
| 50  | 10.0 | 25   | 0.04 | 4/8 | 0.76/1 | -0.58 | Limit cycle† |
| 50  | 10.0 | 50   | 0.17 | 3   | 0.35   | 0.31  | Chaos        |
| 50  | 10.0 | 75   | 0.20 | 3   | 0.28   | 0.44  | Chaos        |
| 100 | 5.0  | 25   | 0.21 | 3   | 0.51   | 0.05  | Chaos‡       |
| 100 | 5.0  | 50   | 0.41 | 5   | 0.53   | 0.38  | Chaos        |
| 100 | 5.0  | 75   | 0.51 | 6   | 0.22   | 0.65  | Chaos        |
| 100 | 10.0 | 25   | 0.02 | 4   | 1      | -0.57 | Limit cycle  |
| 100 | 10.0 | 50   | 0.13 | 3   | 1      | -0.29 | Limit cycle  |
| 100 | 10.0 | 75   | 0.11 | 3   | 0.99   | -0.16 | Limit cycle§ |

Summary of dynamical behaviours of the model with parameter values estimated below. Findings of chaos were confirmed by numerically calculating the dominant Lyapunov exponent<sup>25</sup>. Variables:  $q$ ,  $D$  and  $K'$  are model parameters (described in the text),  $s$  is the standard deviation of log-transformed vole densities (a measure of the amplitude),  $T$  is the lag at which the autocorrelation function (ACF) reached its peak (a measure of the period), ACF[T] is the value of ACF at that period (an indication of the strength of periodicity), and LE1 is the estimated dominant Lyapunov exponent, which measures sensitivity of dependence on initial conditions (chaos indicated by positive values). Parameter estimation. Parameters were estimated as follows, using data primarily from the Alajoki study area in western Finland, where microtine rodents and their predators have been studied since 1977<sup>8,26</sup>.  $r$ : maximal finite rates of increase from spring to autumn occurred in 1981 ( $\times 5.8$ ), 1984 ( $\times 19.5$ ) and 1990 ( $\times 16.4$ ). We used  $r=5.4$  ( $=\ln_e [15 \times 15]$ ; the unit of time is a year);  $r'$ : the maximal finite rates of increase from autumn to spring occurred in 1987/88 ( $\times 11.8$ ) and 1990/91 ( $\times 2.4$ ). We assumed  $r'=r/2$ , thus  $r'=2.7$ ;  $K$ : the maximum mean density of *Microtus* voles in fields in Alajoki has been about 100 individuals per hectare ( $\text{ind. ha}^{-1}$ ) in 1977–1990 (ref. 26), hence  $K=100$ ;  $c$ : least weasel consumes 0.6 g food per g of body mass per day<sup>27</sup>. The mean body masses of male and female least weasels are 48 and 35 g, and the body mass of *Microtus* rodents is 25 g (ref. 8). Additionally, a lactating female mustelid consumes three times the amount of food needed by non-lactating females<sup>28</sup>, and 'surplus killing' may amount to 50% of food requirement<sup>29</sup>. Based on these values, we estimated  $c=600$ ;  $D$ : we are not aware of any data to estimate this parameter, hence we repeated simulations with 3 values that cover the realistic range,  $D=2.5$ , 5 and 10 ( $\text{ind. ha}^{-1}$ );  $v$ : the maximum finite rate of increase of least weasels occurred in summer 1985, when the density increased from 1.3  $\text{ind. km}^{-2}$  in spring, to 5.0 in autumn<sup>8</sup>, which matches the theoretical maximum. Thus  $v=2.8$ ;  $q$ : this is another difficult parameter to estimate, and we repeated simulations with 3 values that should cover the range of realistic values,  $q=50$ , 100 and 200;  $N_{\text{crit}}$ : in 3 years the density of least weasels did not increase in the course of summer, presumably because in these summers the density of rodents was too low to allow breeding by least weasels. The mean spring density of *Microtus* rodents in these years was 2  $\text{ind. ha}^{-1}$ . Others have reported substantially higher critical prey densities for breeding by weasels, 10–14 breeding field voles per  $\text{ha}^{14,15}$ . It is also likely that the value of  $N_{\text{crit}}$  is related to the value of  $D$ . We assumed that  $N_{\text{crit}}=2D$ , thus the  $N_{\text{crit}}$  values used range from 5 to 20;  $d_{\text{high}}$ : maximum mortality of least weasels occurred in winter 1989/90, when the autumn density was 5.0  $\text{km}^{-2}$  and the spring density was 0.3 (ref. 8). We assumed that  $d_{\text{high}}=-5$ ;  $d_{\text{low}}$ : in 3 of 7 winters the density of least weasels apparently increased from autumn to spring. As there is no evidence for winter breeding by least weasels in Fennoscandia, these increases are probably due to movements of weasels. In any case, when prey is abundant, mortality appears to be very low, and we assumed  $d_{\text{low}}=-0.1$ .

\* High-amplitude oscillations similar to dynamics characterizing  $D=2.5$ .

† 8-point limit cycle sometimes dominated by  $T=4$  ( $\text{ACF}[4]=0.76$ ), sometimes by  $T=8$  ( $\text{ACF}[8]=1$ ), depending on initial conditions.

‡ For some initial values, dynamics settle onto a 9-point limit cycle.

§ 18-point or 12-point limit cycle dominated by  $T=3$ .

parameter combinations. We emphasize that all decisions about which values of parameters to include in the numerical explorations of the model were made before we saw the results (parameters estimated by I.H.; model output analysed by P.T.).

Ignoring seasonal variation in numbers, the dynamics predicted by the model range from stability to chaotic oscillations characterized by unrealistically large amplitude (7 orders of magnitude) and long intervals between rodent 'peaks' (> 10 yr). The dynamical extremes are associated with extreme values of the parameters that were varied. Stability was associated with  $q=200$ , whereas high-amplitude oscillations were associated with  $D=2.5$ . Excluding parameter combinations involving these values, the model predicts dynamics that are qualitatively similar to the dynamics of boreal Fennoscandian rodents (Table 1). Figure 1 shows the predicted dynamics for the median parameter values,  $q=100$ ,  $D=5$  and  $K'=50$ ; Figure 2 gives examples of observed dynamics from Fennoscandia.

A more detailed, quantitative comparison between predicted and observed rodent dynamics was provided by analysing long-term data sets from eight Fennoscandian localities, including the

longest series available (Table 2). The data document fluctuations in the pooled numbers of all microtine rodents at a locality as measured by trapping in the autumn. We calculated the following four quantities: (1) the standard deviation of log-transformed vole densities,  $s$ , a measure of the amplitude; (2) the lag at which the autocorrelation function (ACF) reaches its peak,  $T$ , a measure of the period of oscillations; (3) the value of ACF at that period,  $ACF[T]$ , which provides an indication of the strength of periodicity; and (4) the estimated Lyapunov exponent,  $LE1$ , which measures the sensitivity of dependence on initial conditions, and allows one to distinguish between stable point equilibria and limit cycles ( $LE1 < 0$ ) from chaos ( $LE1 > 0$ )<sup>10,11</sup>.

The most striking result arising from these comparisons is that the majority of model-predicted dynamics are chaotic (Table 1), consistent with the results of nonlinear time-series analysis of the real data (Table 2). Only the most southern locality in Table 2 showed non-chaotic, stable dynamics. (This agrees with previous results demonstrating a geographical gradient from high-amplitude, relatively regular oscillations in the north, to low-

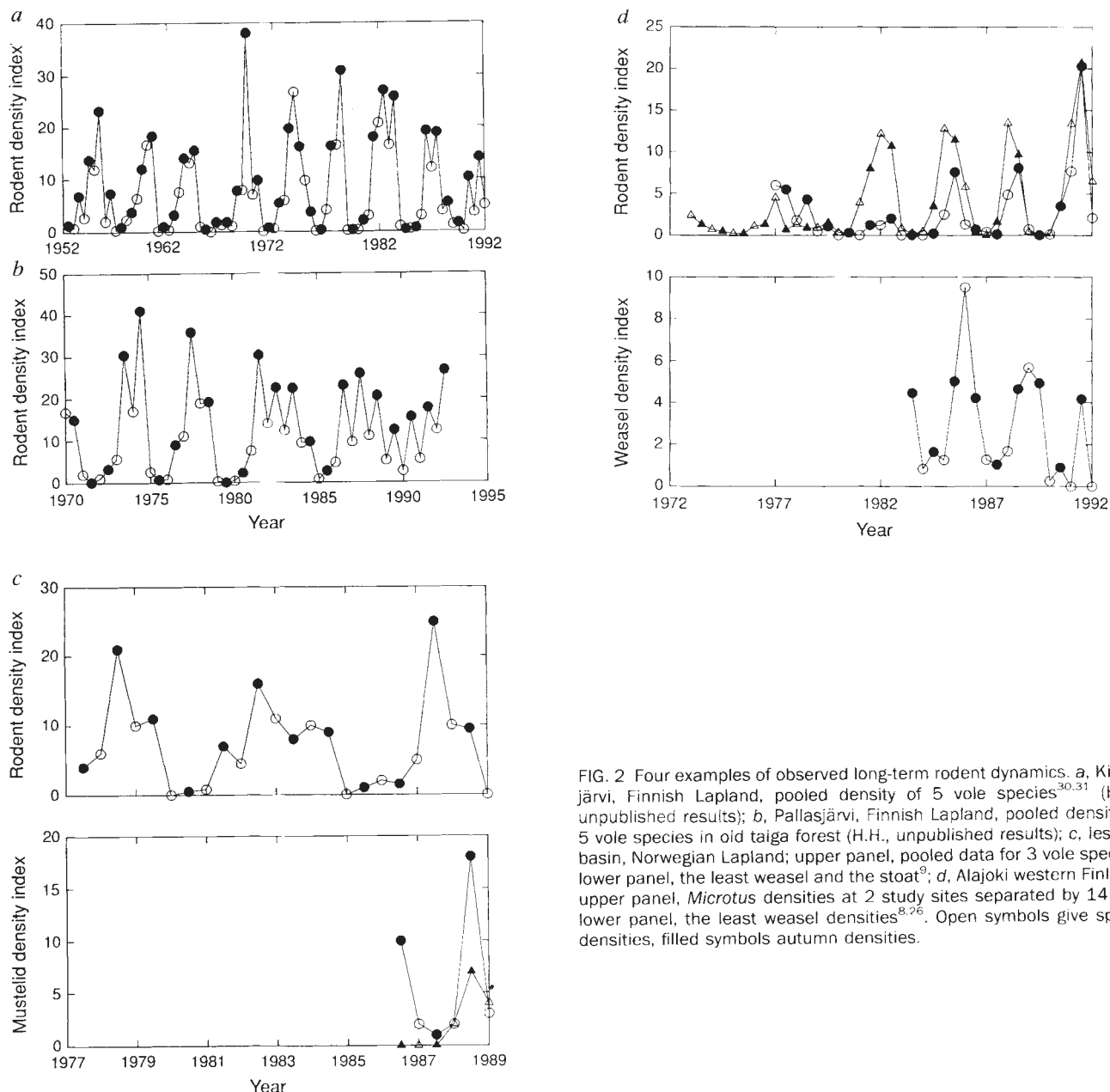


FIG. 2 Four examples of observed long-term rodent dynamics. *a*, Kilpisjärvi, Finnish Lapland, pooled density of 5 vole species<sup>30,31</sup> (H.H., unpublished results); *b*, Pallasjärvi, Finnish Lapland, pooled density of 5 vole species in old taiga forest (H.H., unpublished results); *c*, Iesjavri basin, Norwegian Lapland; upper panel, pooled data for 3 vole species; lower panel, the least weasel and the stoat<sup>9</sup>; *d*, Alajoki western Finland; upper panel, *Microtus* densities at 2 study sites separated by 14 km; lower panel, the least weasel densities<sup>8,26</sup>. Open symbols give spring densities, filled symbols autumn densities.



TABLE 2 Analysis of vole time series

| Locality     | Latitude | Years     | s    | T | ACF[T] | LE1   |
|--------------|----------|-----------|------|---|--------|-------|
| Kilpisjärvi* | 69°      | 1949–1970 | 0.58 | 5 | 0.24   | 0.12  |
| Kilpisjärvi* | 69°      | 1971–1992 | 0.68 | 5 | 0.44   | 0.12  |
| Pallasjärvi  | 68°      | 1970–1992 | 0.73 | 4 | 0.41   | 1.36  |
| Kola         | 67°      | 1946–1964 | 0.79 | 4 | 0.09   | 0.70  |
| Umeå         | 64°      | 1971–1988 | 0.55 | 4 | 0.49   | 0.08  |
| Sotkamo      | 64°      | 1966–1992 | 0.49 | 4 | 0.51   | 0.21  |
| Ruotsala     | 63°      | 1973–1992 | 0.65 | 3 | 0.61   | 0.87  |
| Alajoki      | 63°      | 1977–1992 | 0.77 | 3 | 0.41   | 1.63  |
| Loppi        | 61°      | 1972–1992 | 0.27 | — | —      | –2.06 |

\* Kilpisjärvi data were analysed in two separate pieces to compensate for non-stationarity in this long data set. Variables: s, T, ACF[T] and LE1 as in Table 1.

amplitude, irregular oscillations in the south<sup>4</sup>, probably reflecting the stabilizing effect of abundant generalist predators in the south<sup>7</sup>.) Despite being chaotic, model-predicted oscillations were characterized by statistical periodicities, with a typical period of 3 to 5 yr, although for some more extreme parameter combinations the predicted period was longer. These periodicities were rather weak in strength, with ACF[T] varying from 0.2 to 0.5. The observed oscillations were also characterized by periodicities of 3 to 5 yr, and their strength varied from 0.1 to 0.6. The amplitude of the observed vole fluctuations was somewhat greater than the amplitude predicted by the model. This is not unexpected, because the model does not include stochasticity, which would tend to increase the amplitude. According to these quantitative measures, the dynamics predicted by the model with median values of parameters lie squarely in the middle of the observed spectrum of dynamics.

Figure 2 gives the best-documented empirical data from four Fennoscandian localities. We draw attention to apparent long-term changes in the type of dynamics at two of the four localities. At Pallasjärvi, Finnish Lapland (Fig. 2b), the multiannual cyclic component was evident until the mid-1980s, but has been absent since that time. This change in the pattern of dynamics has been associated with the practical absence of the field vole *Microtus agrestis* and the least weasel *Mustela nivalis* from the study area<sup>6</sup> (H.H., unpublished results). It is not clear why these species have become regionally extinct or very scarce, but the lack of multiannual oscillations in the absence of these two key species is consistent with the idea that the relatively regular oscillations are maintained by their interaction<sup>16</sup>, and become reflected in the dynamics of the other small mammal species because of shared predators<sup>6,16</sup>. In Alajoki, western Finland, four distinct 3-yr cycles have been observed since 1984, but in 1973–1983, rodent density was constantly low (Fig. 2d). Although such irregularities could be expected as a part of strongly chaotic dynamics, here they are more likely to reflect subtle changes in the relative abundances of the various species in the small mammal and predator communities, induced for example by changes in the landscape structure. Our modelling results indicate that biologically plausible adjustments in parameter values may push the dynamics from annual cycles (multiannual stability) to multiannual cycles with long period and high amplitude.

Our results add a critical piece to the growing evidence that the 3–5 yr small mammal cycle in Fennoscandia and probably elsewhere is generated by delayed density dependence as a result of specialist predators. No other hypothesis about the rodent dynamics has been formulated as a quantitative model and been successfully tested with field data. □

- Batzli, G. O. in *Wildlife 2001: Populations* (eds McCullough, D. M. & Barrett, R. H.) 831–850 (Elsevier, 1992).
- Henttonen, H., Oksanen, T., Jortikka, A. & Haukioja, V. *Oikos* **50**, 353–365 (1987).
- Hanski, I., Hansson, L. & Henttonen, H. *J. Anim. Ecol.* **60**, 353–367 (1991).
- Korpimäki, E., Norrdahl, K. & Rinta-Jaskari, T. *Oecologia* **88**, 552–561 (1991).
- Oksanen, L. & Oksanen, T. *Ecography* **15**, 226–236 (1992).
- Turchin, P. *Oikos* (in the press).
- Turchin, P. & Millstein, J. A. *EcoDyn/RSM: Response Surface Modelling of Nonlinear Ecological Dynamics* (Applied Biomathematics, Setauket, NY, 1993).
- May, R. M. *Complexity and Stability of Model Ecosystems* (Princeton Univ. Press, NJ, 1973).
- Hansson, L. & Henttonen, H. *Ann. Zool. Fennici*, **22**, 277–288 (1985).
- Erlinge, S. *Oikos* **25**, 308–314 (1974).
- Tapner, S. J. *Anim. Ecol.* **48**, 603–617 (1979).
- Henttonen, H. *Oikos* **50**, 366–370 (1987).
- Korpimäki, E. & Norrdahl, K. *Oikos*, **55**, 205–215 (1989).
- Hanski, I. *Trends Ecol. Evol.* **6**, 141–142 (1991).
- Erlinge, S. *Oikos* **26**, 378–384 (1975).
- Tapner, S. C. *J. Zool.* **179**, 219–224 (1976).
- Hansson, L. in *Winter Ecology of Small Mammals* (ed. Merritt J. F.) 225–234 (Spec. Publ. Carnegie Mus. Nat. Hist. No. 10, Pittsburgh, 1984).
- Kaikusalo, A. & Tasi, J. in *Winter Ecology of Small Mammals* (ed. Merritt J. F.) 243–252 (Spec. Publ. Carnegie Mus. Nat. Hist. No. 10, Pittsburgh, 1984).
- May, R. M. *Science* **186**, 645–647 (1974).
- May, R. M. *Nature* **261**, 459–467 (1976).
- Wolff, A., Swift, J. B., Swinney, H. L. & Vastano, J. A. *Physica* **16D**, 285–317 (1985).
- Korpimäki, E. & Norrdahl, K. *Oikos* **62**, 195–208 (1991).
- Gillingham, B. J. *J. Mammal.* **65**, 517–519 (1984).
- Erlinge, S. *et al.* *Oikos* **40**, 36–52 (1983).
- Jedrzejska, B. & Jedrzejski, W. *Acta Theriol.* **34**, 347–359 (1989).
- Laine, K. & Henttonen, H. *Oikos* **40**, 407–418 (1983).
- Laine, K. & Henttonen, H. *Oikos*, **50**, 389–395 (1987).

ACKNOWLEDGEMENTS. We thank A. Berryman, M. Gyllenberg, R. May, H. Metz, K. Norrdahl and C. Thomas for discussion and for comments on the manuscript.

## Unilateral neglect restricted to visual imagery

C. Guariglia\*†, A. Padovani‡, P. Pantano‡ & L. Pizzamiglio\*†§

Departments of \*Psychology and ‡Neurology, University of Rome 'La Sapienza', Via dei Marsi 78, 00185 Rome, Italy

† Clinica S. Lucia, Via Ardeatina 306, 00179 Rome, Italy

**DISORDERS** in perceiving and exploring the visual space contralateral to a brain lesion have been frequently described. Many patients with hemi-neglect for extrapersonal space also show neglect in a representational domain when the task requires imagining a well-known piazza from a given vantage point<sup>1,2</sup> or comparing two visual images<sup>3–5</sup>. Cognitive<sup>6</sup> and psychophysiological<sup>7</sup> studies show a functional parallelism between the perceptual and imaginative domain<sup>7</sup>, indicating that spatial perception and imagery share the same neural substrata. Here we describe a patient with a persistent disorder in visual imagery for familiar piazzas in the absence of any neglect for stimuli located in a far<sup>8</sup> or near<sup>9</sup> space or on his own body<sup>10</sup>. Contrary to previous cases involving imagery disorders, computerized tomography scans showed a lesion confined to the right frontal lobe, suggesting the role of the frontal lobe in some specific types of mental imagery.

A 59-year-old right-handed man was examined 16 months after an ischaemic stroke, which produced a left hemiparesis. A computerized tomography (CT) scan showed a large lesion involving the right frontal and anterior temporal lobe (Fig. 1). A SPECT study showed a hypoperfusion in the lesioned frontal area, but not a perfusional deficit remote from the infarct, except in the contralateral cerebellum. A neuropsychological examination showed no mental deterioration or visual field defects, some difficulty in abstract reasoning and face recognition, and a mild impairment in selective attention and in constructional apraxia (Table 1).

The patient's performances on tests for visuo-spatial (Line and Letter Cancellation, Wundt–Jastrow Area Illusion and Sentence Reading) and personal neglect<sup>11</sup> were normal. In contrast, when

Received 22 March; accepted 27 April 1993.

- Elton, C. *Voles, Mice and Lemmings: Problems in Population Dynamics* (Oxford Univ. Press, 1942).
- Krebs, C. J. & Myers, J. H. *Adv. ecol. Res.* **8**, 267–399 (1974).
- Christian, C. J. in *Populations of Small Mammals under Natural Conditions* (ed. Snyder, D. P.) 143–158 (Pymatuning Lab. Ecol. Spec. Publ. 5, 1978).
- Hansson, L. & Henttonen, H. *Trends Ecol. Evol.* **3**, 195–200 (1988).

§ To whom correspondence should be addressed.

he had to describe different but very familiar piazzas from a given vantage point, he consistently reported several elements on the right side and none or very few on the left side (Fig. 2). When requested to describe these piazzas from the opposite perspective, he again reported more elements on the right side of his mental image. The patient's performances were very similar when retested at one- and two-month intervals.

Interestingly, the patient misplaced some semantically relevant details from the left to the right side of one piazza; that is, the ticket office where he had taken a bus to work for 20 years was described as being on the non-neglected side.

Another set of tests was done to identify possible selected forms of neglect (Table 1). The patient showed no disorder of perceptual or motor neglect<sup>12</sup>, or asymmetry in perceiving targets in a far or near space. On tasks involving the generation and manipulation of mental images, he performed like normal controls in comparing mental representations of both colour and weight of objects<sup>13</sup>, and performed normally in a test of mental unfolding<sup>14</sup>.

Through mental imagery, the patient was able to identify shapes differing on either the left or right side<sup>3</sup>, with no asymmetry for either side. He could also read and spell a word orally without omitting its left side, as was described for a neglect patient in ref. 15.

Finally, the patient was taken into an unfamiliar room with different objects and told to observe carefully all details in order to describe them afterwards. When asked to describe the room from two opposite perspectives, the patient consistently omitted half of the elements on the left and none on the right side. Conversely, he made no errors in identifying the observed details by multiple choice (choosing between two different armchairs). This performance matched the good verbal but poor spatial imagery of another recently described patient<sup>16</sup>.

The reported observations describe an impressive dissociation between a long-standing deficit in the mental representation of the hemispace contralateral to the lesion without any sign of impairment in exploring and manipulating the corresponding

TABLE 1 Patient's performance in neuropsychological (A), neglect (B) and imaginal (C) tests

|                |                                       | Patient                               | Normal range                           |
|----------------|---------------------------------------|---------------------------------------|----------------------------------------|
| A              | Line orientation                      | 12/30                                 | 15-30                                  |
|                | Street test                           | 6/11                                  | 6-11                                   |
|                | Face recognition test                 | 30/54                                 | 39-54                                  |
|                | Constructional apraxia                | 10/14                                 | 12-14                                  |
|                | Raven's coloured progressive matrices | 16/36                                 | 15° percentile                         |
|                |                                       | Target detected in the left hemispace | Target detected in the right hemispace |
| B <sub>1</sub> | Near neglect (50 cm)                  |                                       |                                        |
|                | Verbal target                         | 19/30                                 | 17/30                                  |
|                | Non-verbal target                     | 24/30                                 | 23/30                                  |
|                | Far neglect (200 cm)                  |                                       |                                        |
|                | Verbal target                         | 24/30                                 | 24/30                                  |
|                | Non-verbal target                     | 27/30                                 | 26/30                                  |
| B <sub>2</sub> | Motor-visual neglect                  | 10/10                                 | 10/10                                  |
| C <sub>1</sub> |                                       | Patient                               | Normal range                           |
|                | Mental representation of colour       | 14/15                                 | 14-15                                  |
|                | Mental representation of weight       | 11/12                                 | 11-12                                  |
| C <sub>2</sub> | Paper unfolding test                  | 8                                     | 6-12                                   |
| C <sub>3</sub> | Abstract shapes                       | Left-side differences                 | Right-side differences                 |
|                |                                       | 34/40                                 | 35/40                                  |

A: Neuropsychological test scores; B<sub>1</sub>: near and far extrapersonal neglect were assessed by projecting verbal and non-verbal strings of targets within a distracting background in near (50 cm) and far (200 cm) space. The patient pointed out the target items with a light pen; a moderate selective attentional deficit emerged without any bias for one side of the space; B<sub>2</sub>: motor and perceptual neglect. A line cancellation test is presented through a mirror so that the left side of the sheet is shown in the right visual field and vice versa. Patients with motor neglect cross lines on the right side of the sheet (which are actually shown in the left visual hemifield), whereas patients with perceptual neglect cross lines in the left side of the sheet (which are shown in the right visual hemifield); C<sub>1</sub>: mental representation of colour and weight of object; both performances are within the normal range<sup>13</sup>, as well as that in C<sub>2</sub>; C<sub>3</sub>: comparison of same or different shapes.

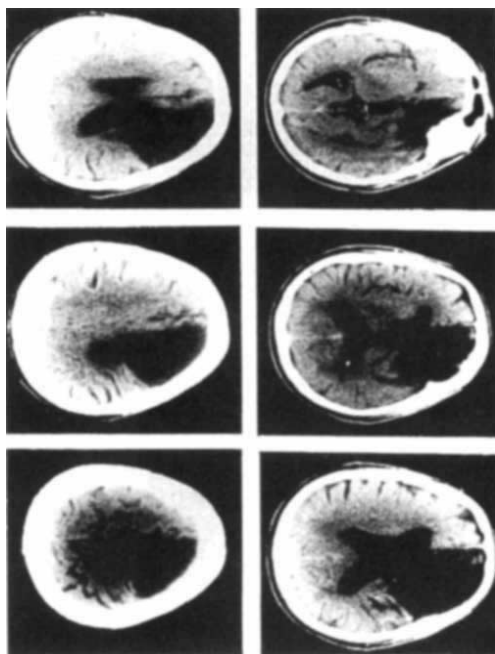


FIG. 1 The CT scan shows a right hemisphere lesion involving the frontal, the anterior temporal lobe, the head of the caudate and lenticular nuclei.

perceptual space and in representing the ipsilateral hemispace. All previously reported representational neglect was observed in patients with personal and/or extrapersonal neglect<sup>1,3,5</sup>. This patient was not affected by a general mental imagery disorder; when comparing mental representations within working memory he performed quite normally (see C<sub>2</sub> and C<sub>3</sub> in Table 1).

These discrepancies of impaired mental imagery could then rely on good access to the immediate memory versus a difficulty in retrieving visuo-spatial information from a long-term store. This interpretation does not entirely correspond with our data as the patient was very competent in tasks relying extensively on long-term acquisitions (comparing colours and weights). The patient seems able to generate the global image of an object or some of its characteristics, but this is insufficient for correctly describing a familiar piazza or a well-inspected room from memory. In this case, the subject has to generate a 'skeletal image' of the square<sup>17</sup> in order to: (1) superimpose a system of egocentric coordinates<sup>18</sup> to organize that space; (2) move an 'attentional window'<sup>17</sup> to select a number of items that belong to the image; and (3) locate the specific items in their respective positions. The latter two operations seem defective in this patient.

The second important finding concerns the predominantly right frontal localization of the lesion. Previous studies<sup>17,19</sup> have



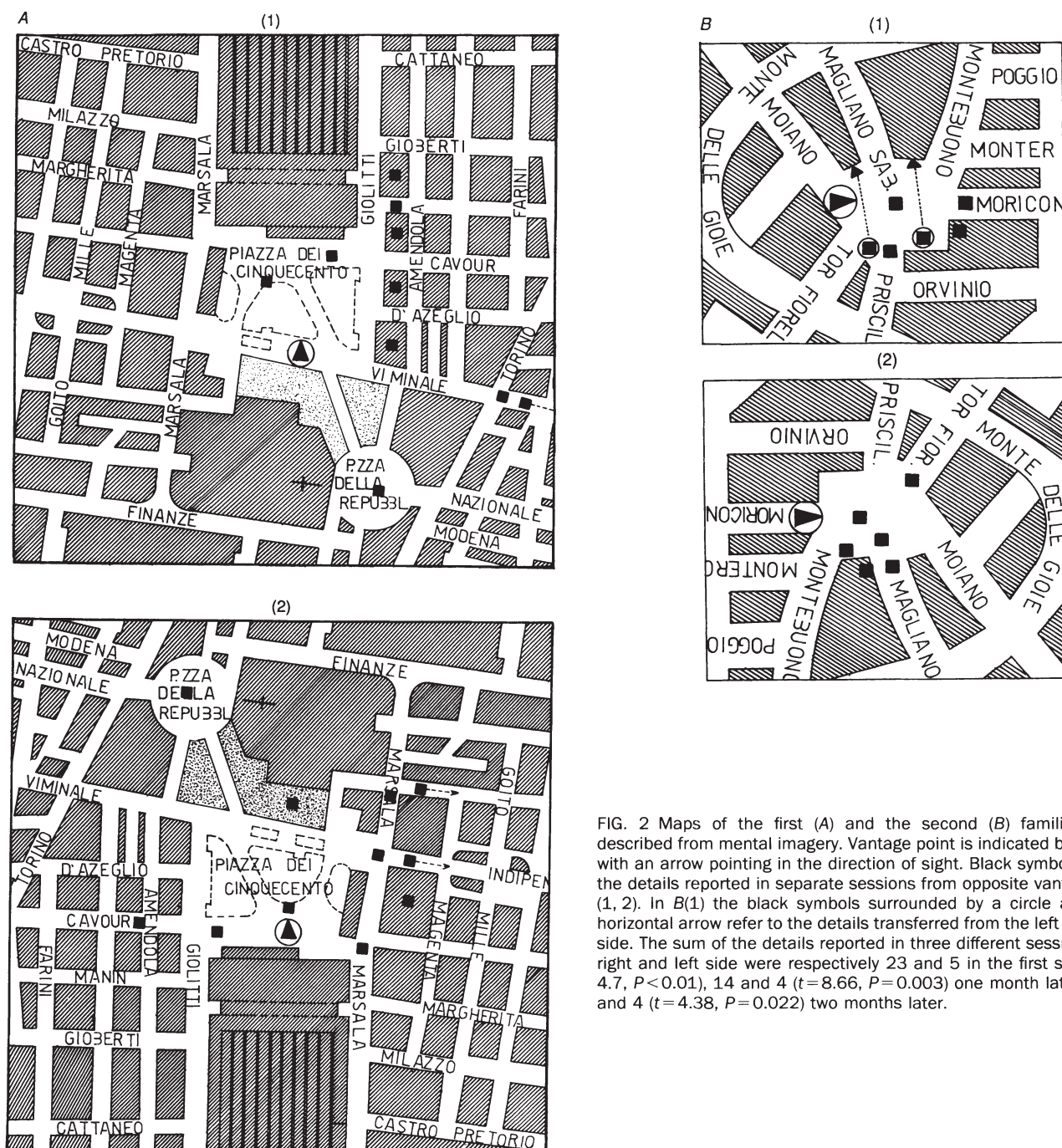


FIG. 2 Maps of the first (A) and the second (B) familiar piazzas described from mental imagery. Vantage point is indicated by the circle with an arrow pointing in the direction of sight. Black symbols indicate the details reported in separate sessions from opposite vantage points (1, 2). In B(1) the black symbols surrounded by a circle and with a horizontal arrow refer to the details transferred from the left to the right side. The sum of the details reported in three different sessions on the right and left side were respectively 23 and 5 in the first session ( $t = 4.7$ ,  $P < 0.01$ ), 14 and 4 ( $t = 8.66$ ,  $P = 0.003$ ) one month later, and 21 and 4 ( $t = 4.38$ ,  $P = 0.022$ ) two months later.

proposed that the neural structures involved in mental imagery were coincident or adjacent to the circuitries involved in visual perception (parieto or temporo-occipital cortex<sup>19</sup>); unlike the present case, no frontal lobe involvement was described.

The tasks used in the psychology literature, however, dealt with the generation and transformation of single objects, symbols or abstract figures, whereas the selective impairment found in the present case only emerged in organizing a complex material from a long-term store into an egocentric frame. Our patient, although showing a good capacity to retrieve many items in the piazzas and in the unfamiliar room from long-term memory, failed when he had to attend to details and position them with respect to the internal spatial coordinates. This matching between a representation of external events and egocentric space may require a new module which shares attentional properties. □

Received 11 December 1992; accepted 16 April 1993.

1. Bisiach, E. & Luzzatti, C. *Cortex* **14**, 129–133 (1978).
2. Bisiach, E. & Berti, A. in *Neurophysiological and Neuropsychological Aspects of Spatial Neglect* (ed. Jeannerod, M.) (North-Holland, Amsterdam, 1987).
3. Bisiach, E., Luzzatti, C. & Perani, D. *Brain* **102**, 609–618 (1979).
4. Grossi, D., Modafferi, A., Pelosi, L. & Trojano, L. *Brain Cogn.* **10**, 18–27 (1989).
5. Ogden, J. A. *Neuropsychologia* **23**, 273–277 (1985).
6. Finke, R. A. *Principles of Mental Imagery* (MIT Press, Cambridge, 1989).
7. Farah, M. J. *Cognition* **18**, 245–272 (1984).
8. Rizzolatti, G., Matelli, M. & Pavesi, G. *Brain* **106**, 655–673 (1983).
9. Halligan, P. W. & Marshall, J. C. *Nature* **350**, 498–500 (1991).
10. Guariglia, C. & Antonucci, G. *Neuropsychologia* **30**, 1001–1009 (1992).
11. Zoccolotti, P. & Judica, A. *Neuropsych. Rehabil.* **1**, 33–44 (1991).
12. Tegner, R. & Levander, M. *Brain* **114**, 1943–1950 (1991).
13. De Vries, L. P. *Neuropsychologia* **29**, 1–18 (1991).
14. Ekstrom, R. B., French, J. W., Harman, H. H. & Dermen, D. *Manual for Kit of Factor-referenced Cognitive Tests* (Princeton Educational Testing Service, NJ, 1976).
15. Kinsbourne, M. & Warrington, E. K. *Brain* **25**, 339–344 (1962).
16. Halligan, P. W., Marshall, J. C. & Wade, D. T. *Neuropsych. Rehabil.* **2**, 125–135 (1992).
17. Kosslyn, S. M. *Psych. Rev.* **94**, 148–175 (1987).
18. Pizzamiglio, L., Frasca, R., Guariglia, C., Inccocchia, C. & Antonucci, G. *Cortex* **56**, 551–559 (1990).
19. Farah, M. J., Weisberg, L. L., Monheit, M. & Perronet, F. J. *cogn. Neurosci.* **1**, 302–317 (1989).

# Selective neuroanatomical plasticity and division of labour in the honeybee

Ginger S. Withers\*, Susan E. Fahrbach\*†  
& Gene E. Robinson\*†

\*Neuroscience Program and †Department of Entomology, University of Illinois, Urbana, Illinois 61801, USA

THE mushroom bodies in the protocerebrum are believed to be the structures of the insect brain most closely associated with higher-order sensory integration and learning<sup>1</sup>. *Drosophila melanogaster* mutants with olfactory learning deficits have anatomically abnormal mushroom bodies or altered patterns of gene expression in mushroom body neurons<sup>2–4</sup>. In addition, anatomical reorganization of the mushroom bodies occurs in adult flies<sup>5</sup>, and possibly in adult honeybees<sup>6,7</sup>; disturbance of electrical activity in this region disrupts memory formation in honeybees<sup>8</sup>. Little is known, however, about the relationship of naturally occurring anatomical changes in the mushroom bodies to naturally occurring behavioural plasticity. We now report that age-based division of labour in adult worker honeybees (*Apis mellifera*) is associated with substantial changes in certain brain regions, notably the mushroom bodies. Moreover, these striking changes in brain structure are dependent, not on the age of the bee, but on its foraging experience, thus demonstrating a robust anatomical plasticity associated with complex behaviour in an adult insect.

Adult worker honeybees spend about the first three weeks of their 4–7-week life performing a variety of tasks within the hive, including caring for the queen and brood ('nursing'). They then make a dramatic behavioural transition and begin to forage for food outside the hive<sup>9</sup>. The temporal shift from within-hive tasks to foraging is typically highly stereotyped but is also flexible: bees can respond to changing colony needs by accelerating, retarding or reversing their behavioural development<sup>10</sup>. To determine whether neuroanatomical plasticity is associated with this natural division of labour, volume estimates of all regions of the honeybee brain were made during three distinct phases of adult behavioural development: newly-emerged 1-day-old bees, nurse bees and forager bees (Fig. 1). Unbiased estimates of regional volumes were obtained using a stereological method, Cavalieri's Direct Estimator of Volume<sup>11</sup>. Our analysis revealed that two regions of the bee brain, the olfactory glomeruli and the mushroom bodies, exhibited distinct patterns of volume changes.

Compared with the brains of 1-day-old bees, the brains of nurse bees had significantly larger olfactory glomerular volumes (Fig. 2c). The only other difference between nurses and younger bees was a significant decrease in the volume occupied by the Kenyon cells (Fig. 2d), the intrinsic neuronal population of the mushroom bodies. No differences were detected in either total brain hemispheric volume or protocerebral volume, nor were any differences observed in other brain regions (Fig. 2a, b).

This pattern of overall stability was also characteristic of the forager brain. Again, total brain hemispheric and protocerebral volumes were not significantly different from those of the 1-day-olds or nurses (Fig. 2a, b). The increased olfactory glomerular volume found in the nurse bees was not maintained in the foragers (Fig. 2c), which is surprising because foragers also depend heavily on olfaction<sup>9</sup>.

Striking differences were again found in the volume of the mushroom bodies (Fig. 2d, e). The volume of Kenyon cells was much less in foragers (by 29.2%) than in 1-day-old bees (Fig. 2d). In contrast, the neuropil volume of the mushroom bodies was significantly greater in foragers (by 14.8%) than in 1-day-old bees (Fig. 2e). The mushroom bodies thus seem to undergo

significant reorganization without any change in overall volume. Nurse bees were intermediate in both Kenyon cell and neuropil volume estimates.

As is typical of arthropod nervous systems, the neuronal somata in the bee brain are separated from the neuropilar regions of synaptic contact. Therefore, comparison of these two regions of the mushroom bodies serves as a valuable first approximation of a 'synapse to neuron' ratio (Fig. 3, typical colony). The ratio of mushroom body neuropil to Kenyon cell body volume was nearly doubled during the transition from inexperienced 1-day-old bee to nurse bee and ultimately to forager (1-day-old bees, 1.3; nurse bees, 1.7; foragers, 2.2). Although we have not directly quantified neuronal and synaptic density in the brains of these bees, studies of vertebrate brains have shown consistently that an increase in volume reflects the addition of synapses<sup>12–14</sup>. Additional neuropil volume in the

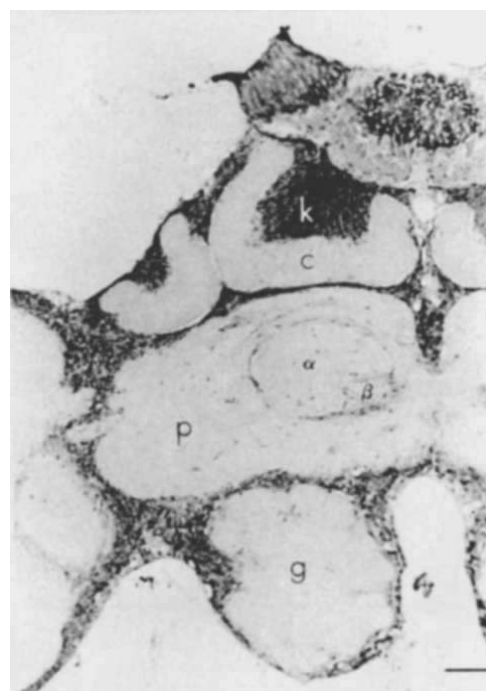


FIG. 1 Representative coronal (frontal) section of the adult worker bee brain. k, Kenyon cell region; c, calyx of the mushroom body; α, α-lobe, peduncular division; β, β-lobe of the mushroom body; p, accessory protocerebral neuropil; g, olfactory glomeruli. Scale bar, 75 μm.

**METHODS.** One-day-old bees were obtained by placing frames of brood in an incubator (33°C) and collecting bees emerging within a 24-h period. Nurse bees and foragers were collected on the same day from the same colony used to obtain the 1-day-old bees. This 'typical' colony was derived from a naturally mated queen and was composed of about 30,000 bees. Nurses were captured after they were observed feeding brood; foragers were captured at the entrance of the hive carrying pollen loads<sup>16</sup>. After capture, bees were brought to the laboratory and anaesthetized on ice. Brains were rapidly removed from the head capsule, fixed (2% paraformaldehyde, 0.15% picric acid, phosphate buffer) and stored at 4°C. Tissue was cryoprotected in sucrose in phosphate buffer for at least 24 h, embedded in OCT (Tissue Tek, Miles) and frozen. Tissue was sectioned on a cryostat microtome at –20°C (10 or 20 μm, balanced by condition). Sections were stained with 0.13% cresyl violet. Standard nomenclature and neuroanatomical descriptions of Kenyon<sup>24</sup> and Mobbs<sup>15</sup> were used, and both intra- and extra-calycal populations of Kenyon cells were included.

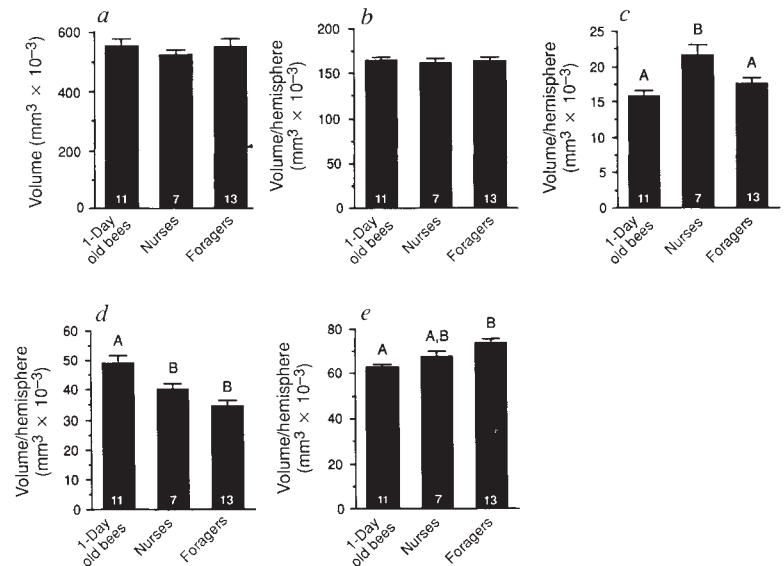


mushroom bodies could be due to a proliferation of the arbors of intrinsic Kenyon cells in the calyces, inputs from other sensory neuropils or efferent cells that originate in the mushroom body lobes<sup>15</sup>.

The observed differences between young bees and foragers may be a consequence of ageing or behavioural state. In the first experiment, foragers were almost certainly 7–10 days older than nurses because they were sampled from a colony with a typical age structure<sup>9</sup>. To dissociate the effects of ageing from the effects of foraging on mushroom body reorganization, we

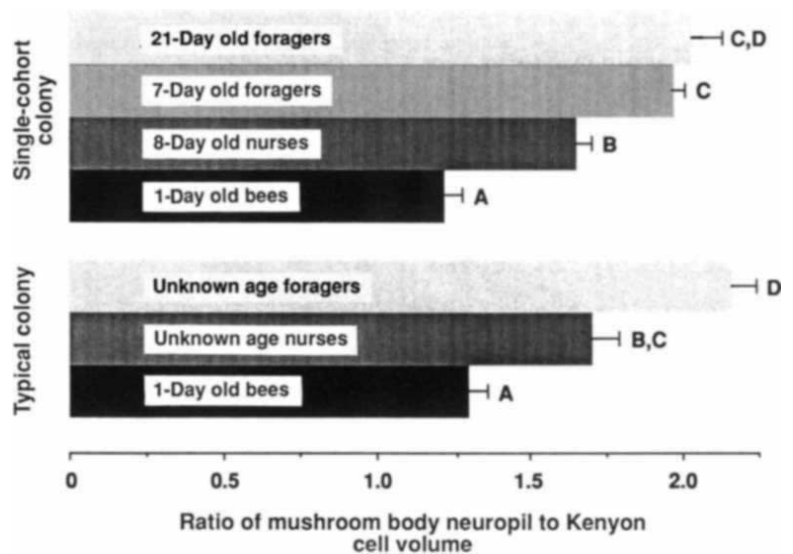
established a 'single-cohort' colony composed initially only of 1-day-old bees<sup>16</sup>. In such a colony, the lack of older foragers accelerates the development of a subset of bees, which are designated 'precocious foragers'<sup>17</sup>. These bees can begin to forage when they are as young as 4 days old, instead of at the typical 21–24 days. Precocious foragers were most similar to foragers of normal age in both Kenyon cell and mushroom body neuropil volumes, again yielding a ratio of neuropil to Kenyon cell body volume of ~2 (Fig. 3, single-cohort colony). As in the first experiment, the volume estimates for the nurse bees were

FIG. 2 Mean brain region volume ( $\text{mm}^3 \times 10^{-3} \pm \text{s.e.}$ ) for bees from three behavioural groups: 1-day-old bees (D1), nurse bees (N) and foragers (F). *a*, Hemispheric volume (including optic lobe and suboesophageal ganglion); *b*, protocerebrum (mushroom bodies, accessory protocerebrum and associated neuronal somata); *c*, olfactory glomeruli; *d*, Kenyon cell region; *e*, mushroom body neuropil. One brain hemisphere per bee was analysed. Other regions measured but not significantly different were optic lobe: D1,  $235.6 \pm 22.4$ ; N,  $224.8 \pm 13.9$ ; F,  $246.4 \pm 20.8$ ; accessory protocerebrum: D1,  $53.7 \pm 2.8$ ; N,  $54.7 \pm 4.2$ ; F,  $55.2 \pm 2.6$ ; tritocerebrum and the suboesophageal ganglion: D1,  $70.7 \pm 5.3$ ; N,  $60.3 \pm 8.2$ ; F,  $71.7 \pm 4.4$ . Total mushroom body volume (Kenyon cell region + mushroom body neuropil) also showed no significant change: D1,  $112.2 \pm 3.0$ ; N,  $108.5 \pm 3.1$ ; F,  $108.8 \pm 3.3$ . Data were analysed according to ref. 25 using the General Linear Model and the Student Newman–Kuels test. Bars marked with different letters differed significantly ( $p < 0.05$ ) from each other. The increased volume in the antennal lobe and the mushroom bodies is selective and not merely a reflection of overall brain growth; no task-related differences were detected in either total brain volume or protocerebral volume, nor in any other brain regions. The restricted and bidirectional regional changes also argue against the possibility that the results are an artefact of age-related differential shrinkage of tissues during histological processing. METHODS. Volume was estimated according to ref. 11 using an unbiased stereological method which has been validated by fluid displacement<sup>26</sup>. Camera lucida tracings of 10- or 20- $\mu\text{m}$ -thick sections were made every 80  $\mu\text{m}$ , beginning with a randomly selected section from within the first 80  $\mu\text{m}$  of tissue. Initial investigations based on serial reconstructions indicated that the 80- $\mu\text{m}$  interval was



sufficient to estimate the volume of the brain regions under study, including the mushroom bodies, with a standard error of less than 5%. Area was estimated by a standard stereological point counting system<sup>11</sup>. Volume estimates were obtained by multiplying the area of each section traced by the distance between traced sections (80  $\mu\text{m}$ ). Tracing and volume estimation were done blind to experimental group.

FIG. 3 Mean ratio of mushroom body neuropil to Kenyon cell region. Ratios are shown for the bees from the typical colony depicted in Fig. 2 and for 1-day old bees ( $n = 4$ ), normally aged (8-day-old) nurse bees ( $n = 6$ ), precocious (7-day-old) foragers ('7F',  $n = 5$ ), and normally aged (21-day-old) foragers ('21F',  $n = 4$ ) from a single-cohort colony. There were significant differences in both Kenyon cell volume and mushroom body neuropil volume between 1-day-old bees and: 7-day-old foragers; 21-day-old foragers and 8-day-old nurses. No differences for either region were found between: 7-day-old foragers and 21-day-old foragers; 7-day-old foragers and 8-day-old nurses; and 21-day-old foragers and 8-day-old nurses (Kenyon cell region: D1,  $46.5 \pm 1.6$ ; N,  $40.6 \pm 2.08$ ; 7F,  $36.4 \pm 0.8$ ; 21F,  $37.1 \pm 2.4$ ; mushroom body neuropil: D1,  $57.4 \pm 2.7$ ; N,  $67.2 \pm 4.0$ ; 7F,  $71.6 \pm 0.9$ ; 21F,  $74.8 \pm 2.2$ ). Volume estimates and statistical analyses as in Fig. 2 legend. METHODS. The single-cohort colony was established with 1,536 1-day-old bees (obtained as in Fig. 1 legend), a queen, one frame of honey and pollen and one empty honeycomb frame for the queen to lay eggs in. All bees were marked with a spot of paint on the thorax to ensure that the bees analysed were of known age. Nurses and foragers were collected as in Fig. 1 legend. Precocious foragers had at most 3 days foraging experience. All bees from the single-cohort colony were derived from a queen instrumentally inseminated with the semen of a single drone<sup>27</sup>. Both the queen and drone were unrelated



to the queen from the typical colony. The population of the single-cohort colony was smaller than that of typical colonies, but previous studies have demonstrated that bees in small colonies exhibit normal behaviour<sup>9,10,20</sup>.

intermediate between the values for the 1-day-old bees and foragers, even though these nurses were one day older than the precocious foragers. Age is therefore a less useful predictor of mushroom body organization than is behaviour.

In contrast to the nurse bee, whose behaviour within the dark hive takes place in an intensely social milieu of odour-based communication, a forager navigates the world outside the hive alone, searching for food. Foraging is a demanding task which requires flexible use of both visual and olfactory cues, formation of long-term memories, learning how to handle diverse floral resources, and abstract communication via the dance language<sup>9,18–20</sup>. Such multimodal integration has been proposed as the primary function of the mushroom bodies<sup>1</sup>. In the honeybee, this structure receives major visual and olfactory inputs and provides efferents at least to the central complex and suboesophageal centres. As the calyces of the mushroom bodies are subdivided into regions receiving primarily visual (the 'collar') or primarily olfactory (the 'lip') input as well as a region (the 'basal ring') that receives both<sup>21</sup>, the bee would thus be an excellent model for examining the functional organization of the mushroom bodies. In addition, the observed volume changes in the mushroom bodies of the honeybee brain raise important questions regarding the extent to which neuronal plasticity can anticipate a functional demand and the extent to which experience is required to refine synaptic connectivity.

Results from the brains of precocious foragers imply that earlier task performance is correlated with an earlier assumption of the typical forager brain configuration, but we do not yet know if these changes precede or result from foraging. Because the onset of foraging in honeybees is associated with an increase in the level of juvenile hormone<sup>16</sup>, it will be possible to examine the interaction of endocrine signals with brain structure; studies are underway to determine whether juvenile hormone has direct effects on bee brain structure.

It has long been thought that changes in the strength or pattern of neuronal connections would be necessary for behavioural reorganization<sup>22,23</sup>. Our study of a simple nervous system which generates complex social behaviour offers a new entrée into the cellular mechanisms of behavioural plasticity. □

Received 11 January; accepted 4 May 1993.

- Erber, J., Homberg, U. & Gronenberg, W. in *Arthropod Brain: Its Evolution, Development, Structure and Functions* (ed. Gupta, A. P.) 484–512 (Wiley, New York, 1987).
- Heisenberg, M., Borst, A., Wagner, S. & Byers, D. J. *Neurogenet.* **2**, 1–30 (1985).
- Nighorn, A., Healey, M. J. & Davis, R. L. *Neuron* **6**, 455–467 (1991).
- Han, P.-L., Levin, L. R., Reed, R. R. & Davis, R. L. *Neuron* **9**, 619–627 (1992).
- Technau, G. M. J. *Neurogenet.* **1**, 113–126 (1984).
- Brandon, J. & Coss, R. *Brain Res.* **252**, 51–61 (1982).
- Coss, R., Brandon, J. & Globus, A. *Brain Res.* **192**, 49–59 (1980).
- Erber, J., Masuhr, T. & Menzel, R. *Physiol. Ent.* **5**, 343–358 (1980).
- Winston, M. L. *The Biology of the Honey Bee*, 89–110 (Harvard Univ. Press, Cambridge, 1987).
- Robinson, G. E. A. *Rev. Ent.* **37**, 637–665 (1992).
- Gundersen, H. J. G. et al. *Acta path. microbiol. Immun. scand.* **96**, 379–394 (1988).
- Black, J. E., Isaacs, K. R., Anderson, B. J., Alcantara, A. A. & Greenough, W. T. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5568–5572 (1990).
- DeVoogd, R. J., Nixdorf, B. & Nottebohm, F. *Brain Res.* **329**, 304–308 (1985).
- Turner, A. M. & Greenough, W. T. *Brain Res.* **329**, 195–203 (1985).
- Mobbs, P. G. *Phil. Trans. R. Soc. Lond.* **B298**, 309–354 (1982).
- Robinson, G. E., Page Jr, R. E., Strambi, C. & Strambi, A. *Science*, **246**, 109–112 (1989).
- Huang, Z.-Y. & Robinson, G. E. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11726–11729 (1993).
- Menzel, R. in *Neurobiology of Comparative Cognition* (eds Kesner, R. P. & Olton, D. S.) 237–292 (Lawrence Erlbaum Associates, Hillsdale, 1990).
- Gould, J. L. A. *Rev. Psychol.* **37**, 163–192 (1986).
- von Frisch, K. *The Dance Language and Orientation of Bees* (Belknap/Harvard Univ. Press, Cambridge, 1967).
- Mobbs, P. G. J. *Insect Physiol.* **30**, 43–58 (1984).
- Ramón y Cajal, S. *Proc. R. Soc. Lond.* **55**, 444–468 (1894).
- Tanzi, E. *Riv. sper. Freniat. Med. leg. Alien. ment.*, **19**, 419–472 (1893).
- Kenyon, F. C. J. *comp. Neurol.* **6**, 133–210 (1896).
- SAS Institute SAS Users Guide: Statistics Version 5 (SAS Institute, Cary, 1985).
- Michel, R. P. & Cruz-Orive, L. M. J. *Microsc.* **150**, 117–136 (1988).
- Laidlaw, H. H. *Instrumental Insemination of Honey Bee Queens* (Dadant, Hamilton, 1977).

ACKNOWLEDGEMENTS. We thank Z.-Y. Huang, J. C. Kuehn, S. D. Schultz and D. Nichols for technical assistance, and T. Giray, W. T. Greenough, Z.-Y. Huang, M. E. O'Neal, S. T. Trumbo and C. S. Wallace for helpful comments on the manuscript. This work was supported by grants from the NIH (S.E.F.), NIMH (G.E.R.) and NSF (S.E.F. and G.E.R.).

## Voltage-dependent potentiation of L-type $\text{Ca}^{2+}$ channels due to phosphorylation by cAMP-dependent protein kinase

Adrian Sculptoreanu, Todd Scheuer & William A. Catterall

Department of Pharmacology, SJ-30, University of Washington, Seattle, Washington 98195, USA

THE force of contraction of motor units in skeletal muscle is graded by changing the discharge rate of motor neurons<sup>1</sup>, and cytosolic calcium transients are similarly increased<sup>2</sup>. During single twitches, contraction is not dependent on extracellular calcium<sup>3</sup>, and L-type  $\text{Ca}^{2+}$  channels may only function as voltage sensors for initiating  $\text{Ca}^{2+}$  release from the sarcoplasmic reticulum<sup>4–6</sup>. In contrast, forceful tetanic contractions triggered by action potentials at high frequency (20 to 200 Hz) are dependent on extracellular  $\text{Ca}^{2+}$  concentration and sensitive to L-type  $\text{Ca}^{2+}$  channel antagonists<sup>7–9</sup>, but the mechanism of regulation of contractile force is unknown. Here we report a large, voltage- and frequency-dependent potentiation of skeletal muscle L-type  $\text{Ca}^{2+}$  currents by trains of high-frequency depolarizing prepulses, which is caused by a shift in the voltage-dependence of channel activation to more negative membrane potentials and requires phosphorylation by cyclic AMP-dependent protein kinase in a voltage-dependent manner. This potentiation would substantially increase  $\text{Ca}^{2+}$  influx and contractile force in skeletal muscle fibres in response to tetanic stimuli.

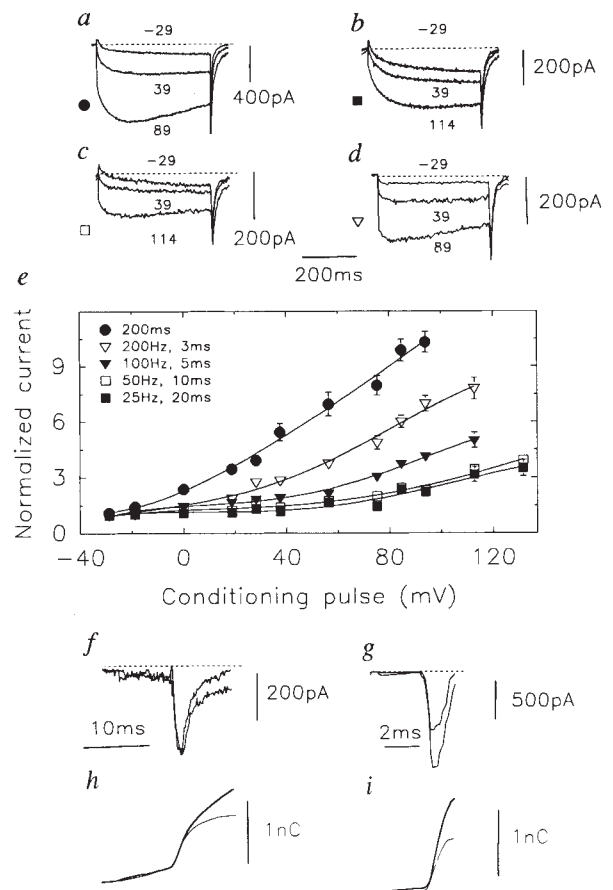
L-type  $\text{Ca}^{2+}$  currents recorded during depolarizations to voltages just beyond threshold (–29 mV, for example) in the whole-cell voltage-clamp configuration in rat skeletal muscle myoballs were unaffected by a prepulse to –29 mV, but were strongly potentiated when preceded by a single conditioning depolarization in the range of 0 to +89 mV for 200 ms (Fig. 1a and e; filled circles). Prepulse trains (20 ms at 25 Hz, 10 ms at 50 Hz, 5 ms at 100 Hz, or 3 ms at 200 Hz) to +39 mV, +89 mV or +114 mV potentiated the  $\text{Ca}^{2+}$  currents activated at –29 mV (Fig. 1b–d). Brief, rapid stimuli like those occurring during muscle tetani were more effective than longer, less frequent stimuli. At +39 mV, near the peak of the action potential, stimulation at 200 Hz with 3-ms pulses increased the peak  $\text{Ca}^{2+}$  current nearly 3-fold; more positive conditioning stimuli produced substantially stronger potentiation (Fig. 1e).

After potentiation,  $\text{Ca}^{2+}$  currents activated more rapidly upon depolarization and deactivated more slowly upon repolarization (Fig. 1a–d). Because of the much greater driving force for  $\text{Ca}^{2+}$  entry after repolarization, integrated  $\text{Ca}^{2+}$  influx during repolarization accounts for most of the  $\text{Ca}^{2+}$  entry during each pulse cycle (Fig. 1f–i). It is substantially increased by prepulses to +39 mV.  $\text{Ca}^{2+}$  entering the cytosol upon repolarization could enhance contraction directly or be available for release from the sarcoplasmic reticulum during the subsequent action potential, leading to a progressive increase in contractile force during tetanic stimulation.

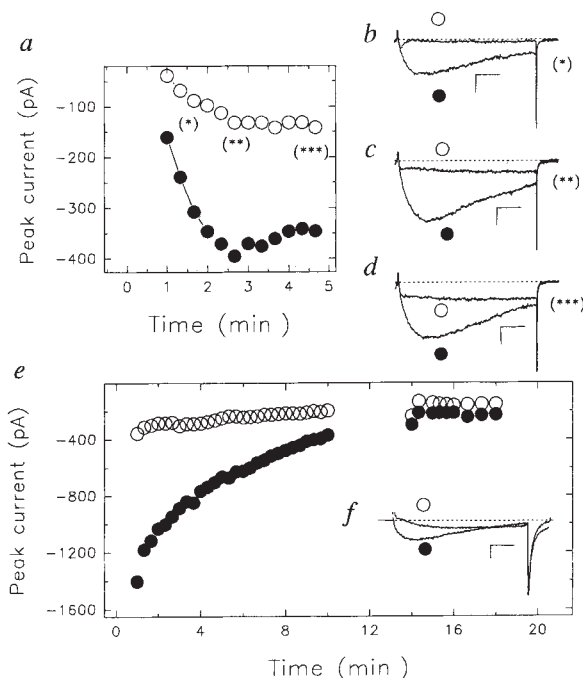
$\text{Ba}^{2+}$  currents through L-type  $\text{Ca}^{2+}$  channels were also strongly potentiated when preceded by single conditioning depolarizations to +47 mV (Fig. 2a–d). Prepulse-induced potentiation was reversible in the 20 s between depolarizations. Soon after breaking into the cell, there was little L-type  $\text{Ba}^{2+}$  current during test pulses to –19 mV, but it increased 5- to 10-fold after a conditioning prepulse (Fig. 2b). The increased current was entirely blocked by the L-type  $\text{Ca}^{2+}$  channel antagonist PN200-110 (data not shown). As the experiments progressed, currents measured without a prepulse increased steadily (Fig. 2a, c, d; open circles) owing to a negative shift of



**FIG. 1** Frequency and voltage-dependent potentiation of L-type  $\text{Ca}^{2+}$  currents. After establishing the whole-cell configuration, currents during test depolarizations were monitored until they reached a steady magnitude (Fig. 2a). The following protocol was then applied once every 20 s. An initial 300-ms-long test depolarization to  $-29$  mV was followed 6 s later by a conditioning procedure, a 20-ms period at  $-56$  mV and then a second test pulse identical to the first. The conditioning procedure was either a single 200-ms conditioning prepulse or a 200-ms train of conditioning depolarizations. *a–d*, Currents recorded during the test depolarization following single 200-ms-long conditioning depolarizations to the indicated potentials (*a*), or following 200-ms conditioning trains of 20-ms pulses applied at 25 Hz (*b*), 10-ms pulses applied at 50 Hz (*c*) or 3-ms pulses applied at 200 Hz (*d*). *e*, Peak inward current recorded during the test depolarization following the conditioning described in *a–d* was normalized to the current recorded during the preceding test depolarization and plotted versus the potential of the conditioning depolarizations. *f*, Time course of  $\text{Ca}^{2+}$  current during the first and last 10-ms depolarization to  $+39$  mV and repolarization to  $-56$  mV in a 200-ms, 50 Hz train. *g*, Time course of the  $\text{Ca}^{2+}$  current during the first and last 3-ms depolarization to  $+39$  mV and repolarization to  $-56$  mV in a 200-ms 200 Hz train. *h*, *i*, Cumulative integrals of the  $\text{Ca}^{2+}$  current traces shown in *f* and *g*, respectively. The lighter, lower trace corresponds to the first depolarization and repolarization of the train, and the heavier higher trace to the last depolarization and repolarization of the train. **METHODS.** Skeletal muscle myoblasts were prepared from 20-day-old rat embryos as described<sup>19</sup>. After 5 days, cultures were treated with  $10\ \mu\text{M}$  colchicine to form myoballs 20 to 100  $\mu\text{m}$  in diameter which were viable for electrophysiological experiments for two to three weeks. The pipette (intracellular) solution for whole-cell recording contained (in mM): *N*-methyl-D-glucamine (130), ATP (0.3) (free-acid), EGTA (20) (free acid), BAPTA (5), HEPES (10),  $\text{Mg}(\text{OH})_2$  (6), and  $\text{Ca}(\text{OH})_2$  (4), with pH adjusted to 7.3 with methanesulphonate. For the experiments shown, the extracellular solution contained (in mM): calcium methylsulphonate (2), tetraethylammonium methylsulphonate (120), 4-aminopyridine (5), HEPES (10), pH 7.4. All experiments were done at room temperature ( $20\text{--}23\ ^\circ\text{C}$ ). Under these conditions, net outward delayed rectifying currents were predominant over a wide voltage range immediately after establishing whole-cell configuration. Within 5–10 min, these outward currents disappeared, presumably as a result of dialysis of the cell with impermeant ions and  $\text{Ca}^{2+}$  chelators. After this time, inward currents carried by L-type  $\text{Ca}^{2+}$  channels were observed without contamination from outward currents. The extracellular solution for experiments in subsequent figures contained (in mM): barium methanesulphonate (20), tetraethylammonium methanesulphonate (90), 4-aminopyridine (5), HEPES (10), pH 7.4. No outward currents were observed after establishing the



whole-cell configuration in this solution. Whole-cell currents<sup>20</sup> were recorded using a List EPC-7 patch clamp amplifier. Currents were filtered at 2 kHz (8-pole Bessel filter;  $-3\ \text{dB}$ ). Pulse generation, current recording and data analysis were performed using software based on the Fastlab system (Indec Systems). Peptide inhibitors of cAPK (PKI 5-24, Peninsula Labs) and of protein kinase C were maintained as 10 mM aqueous stocks. The pseudosubstrate site inhibitor peptide of PKC<sup>21</sup> was a gift from K. Meier and E. G. Krebs.



**FIG. 2** Time- and voltage-dependent potentiation of L-type  $\text{Ca}^{2+}$  channel currents with  $\text{Ba}^{2+}$  as permeant ion. Starting immediately after establishing the whole-cell configuration, a 3-pulse protocol was repeated once every 20 s. First a 300-ms test depolarization to  $-19$  mV was applied from a holding potential of  $-94$  mV. After returning to the holding potential for 6 s, a 200-ms conditioning prepulse to  $+47$  mV was applied, followed by repolarization to  $-56$  mV for 20 ms. Finally, the same 300-ms test pulse to  $-19$  mV was repeated. *a*, Peak current during the first test pulse (open circles; no conditioning prepulse) and during the second test pulse (filled circles; with conditioning prepulse) as a function of time after breaking into the cell. *b–d*, Test pulse current recordings without (open circles) and with (filled circles) conditioning prepulses immediately on break-in (*b*) and 3 min (*c*) and 5 min (*d*) later. Asterisks in *a* indicate the time at which corresponding current records in *b–d* were recorded. A small inactivating inward current was present soon after break-in in this cell, but disappeared with dialysis (*b*, upper trace). The L-type current activated slowly through the duration of the test pulse and peak currents were measured at the end of the test pulse, so the small inactivating current had no effect on the results. *e*, Same protocol as in *a*, but with PKI in the pipette. *f*, Current records without and with a depolarizing conditioning prepulse after 18 min with PKI in the pipette. Scale bars are 100 pA and 50 ms in all panels.

5–10 mV in the voltage dependence of activation ( $n = 95$  cells). Potentiation of the current by conditioning pulses is caused by a further negative shift in the voltage dependence of activation. With 20 mM  $\text{Ba}^{2+}$  (Fig. 3a) or 2 mM  $\text{Ca}^{2+}$  ( $n=17$ ; not shown), L-type  $\text{Ca}^{2+}$  currents activated at test pulse potentials positive to  $-30$  mV and peaked near 0 mV in the absence of a prepulse. After prepulses to  $+85$  mV, the threshold for current activation shifted to  $-50$  mV, and peak currents were recorded at  $-10$  mV. Potentiation caused a 5- to 10-fold increase in current during test pulses in the range of  $-20$  mV to 0 mV, but no detectable increase at test pulse potentials more positive than  $+20$  mV (Fig. 3a), indicating a negative shift of  $\sim 15$  mV in the voltage dependence of activation.

As the activity of skeletal muscle  $\text{Ca}^{2+}$  channels is enhanced by cAMP-dependent protein kinase (cAPK)<sup>10</sup>, we tested the effect of 10  $\mu\text{M}$  of a specific peptide inhibitor of cAPK (PKI)<sup>11</sup> in the pipette solution. L-type  $\text{Ba}^{2+}$  currents became progressively smaller, and potentiation declined dramatically as PKI diffused into the cell (Fig. 2e and f;  $n = 15$ ). Current-voltage curves were shifted towards more positive potentials compared to controls, and the negative shift in the voltage dependence of activation due to potentiation was blocked (Fig. 3b). These results show that both the spontaneous shifts in the voltage dependence of activation to more negative membrane potentials observed under control conditions and the depolarization-induced potentiation and shift in voltage dependence require cellular cAPK activity.

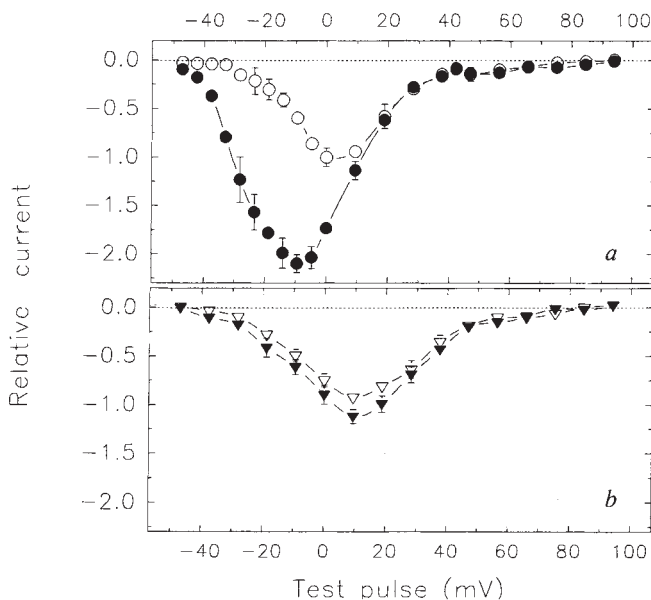


FIG. 3 Effects of conditioning depolarization and cAPK inhibitor on the voltage dependence of  $\text{Ca}^{2+}$  channel activation. Average current-voltage curves from peak test pulse currents obtained without (open symbols) and with (filled symbols) 200-ms-long conditioning depolarizations to  $+85$  mV. **a**, Cells with control pipette solution (circles;  $n=14$ ). **b**, Cells with 10  $\mu\text{M}$  PKI added to the pipette solution (triangles;  $n=10$ ). From a holding potential of  $-95$  mV, the voltage was stepped to  $-56$  mV for 20 ms. For the next 200 ms the membrane was either maintained at  $-56$  mV (no conditioning prepulse) or was stepped to  $+85$  mV (conditioning prepulse). After returning to  $-56$  mV for another 20 ms, a test pulse to a variable potential was applied. Current-voltage curves obtained without conditioning prepulses were normalized to peak inward current. The companion curve obtained with conditioning prepulse was scaled by the same factor. Normalized curves obtained without and with the conditioning prepulse were averaged. The mean inward current densities measured at the peak of the current-voltage relationship without a conditioning prepulse were  $-5 \pm 0.2$   $\mu\text{A}$  per  $\mu\text{F}$  ( $n=24$ ) in control and  $-1.2 \pm 0.6$   $\mu\text{A}$  per  $\mu\text{F}$  ( $n=11$ ) with PKI in the pipette solution.

Voltage-dependent potentiation is observed for conditioning depolarizations to  $-20$  mV and increases to 3.2-fold at  $+37.5$  mV, near the peak of the action potential (Fig. 4). Inclusion of 100 nM PKI (not shown) or 10  $\mu\text{M}$  PKI (Fig. 4a, open triangles) in the pipette solution dramatically inhibited the potentiation at all prepulse voltages. In contrast, adding 1.6  $\mu\text{M}$  of the pseudo-substrate inhibitor of PKC had little effect (Fig. 4a, filled circles). If phosphorylation is required for potentiation, inhibition of phosphoprotein phosphatases should make the pulse-by-pulse potentiation irreversible and cumulative, and should occlude the effect of further conditioning pulses. Inclusion of 10  $\mu\text{M}$  of the phosphatase inhibitor okadaic acid in the recording pipette caused a slow pulse-by-pulse increase in the L-type current recorded in the unconditioned test pulse (Fig. 4e, open symbols), but not in the L-type currents in unstimulated cells (not shown). The current in the conditioned test pulse was nearly unchanged (Fig. 4e, filled symbols). After 10 min of stimulation in the presence of the phosphatase inhibitor, the effect of conditioning prepulses was nearly completely occluded (compare open and filled symbols). Evidently, reversal of potentiation in the period between conditioning pulses depends on phosphoprotein phosphatase activity.

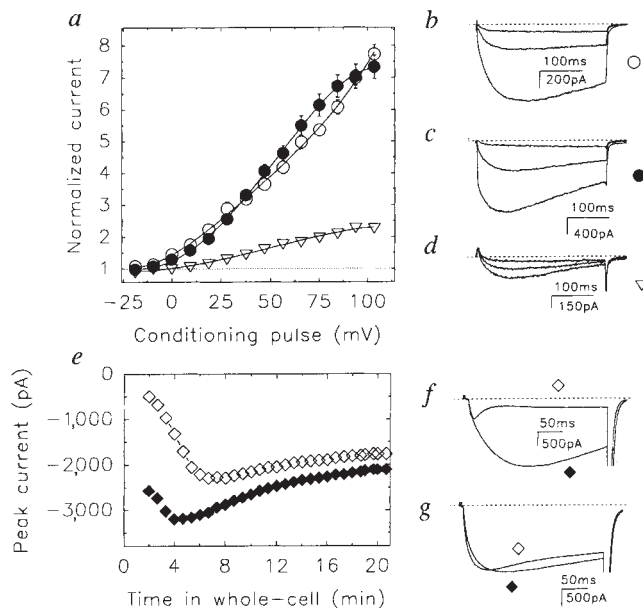


FIG. 4 Effects of inhibitors of protein kinases and phosphoprotein phosphatases on potentiation of  $\text{Ca}^{2+}$  channel activity. **a**, Voltage dependence of potentiation in control experiments (open circles;  $n=21$ ) and after addition of either the peptide inhibitor of protein kinase C<sup>21</sup> (filled circles;  $n=11$ ) or PKI (open triangles;  $n=14$ ) to the pipette solution. The same pulse protocol as in Fig. 2 was used, except that the conditioning pulse potential was varied. Currents measured during the second test pulse to  $-19$  mV were normalized relative to currents measured during the first test pulse in each 3-pulse series. **b–d**, Examples of current traces during test pulses to  $-19$  mV following conditioning pulses to  $-19$ ,  $+47$  and  $+113$  mV. Symbols beside each set of traces refer to the conditions in **a**. **e**, Peak  $\text{Ca}^{2+}$  currents as a function of time after breaking into the cell with 10  $\mu\text{M}$  okadaic acid in the recording pipette before (open symbols) and after (filled symbols) a conditioning pulse to  $+47$  mV. Peak  $\text{Ca}^{2+}$  currents increase during each stimulation pulse protocol in stimulated cells, but do not reverse completely between pulses. This progressive increase occludes the effect of prepulse potentiation. But okadaic acid does not change peak  $\text{Ca}^{2+}$  currents in cells that are not repetitively stimulated, indicating that phosphorylation during the depolarizing prepulse itself causes the increase in peak currents. **f, g**, Example of current records immediately after break-in (**f**) and after 20 min of recording (**g**).



Modulation of  $\text{Ca}^{2+}$  channel activity is often dependent on G proteins. In contrast to the marked effects of reagents that alter cAMP regulatory pathways, inclusion of GDP- $\beta$ -S or GTP- $\gamma$ -S in the recording pipette neither mimicked, inhibited, nor occluded voltage-dependent potentiation. We suggest that this potentiation is due to a voltage-activated, state-dependent phosphorylation of the  $\text{Ca}^{2+}$  channel  $\alpha 1$  or  $\beta$ -subunits by cAPK<sup>6</sup>, in which voltage-dependent phosphorylation results from a voltage-dependent conformational change in the substrate site. The potentiation is reversed between depolarizations by the action of phosphoprotein phosphatases. Evidently, simultaneous depolarization and activation of cAPK are required for this marked potentiation of skeletal muscle  $\text{Ca}^{2+}$  channel activity.

The activity of cardiac L-type  $\text{Ca}^{2+}$  channels is potentiated by depolarizing prepulses<sup>12-14</sup>, and voltage-dependent facilitation requiring phosphorylation by an unidentified kinase activates a quiescent class of L-type  $\text{Ca}^{2+}$  channels in chromaffin cells<sup>15</sup>. However, in studies of cut frog skeletal muscle fibres, prior depolarization accelerated activation of L-type  $\text{Ca}^{2+}$  currents but did not increase peak  $\text{Ca}^{2+}$  currents<sup>16,17</sup>. This may have resulted from a reduction in cAPK activity during the dialysis of cellular contents in the cut fibre preparation. Following inhibition of cAPK, we observe only an acceleration of channel activation without a substantial increase in peak current (Fig. 2f).

Single depolarizations simultaneously activate only a small fraction (<5%) of skeletal muscle  $\text{Ca}^{2+}$  channels<sup>18</sup>. Voltage- and phosphorylation-dependent potentiation may greatly increase the fraction of skeletal muscle  $\text{Ca}^{2+}$  channels that is

activated by repetitive depolarizations. Because sustained, forceful contractions are induced by tetanic trains of action potentials, it is likely that the potentiation of  $\text{Ca}^{2+}$  channel activity described here is induced by these physiological stimuli and causes or contributes to the observed increase in contractile force. □

Received 9 December 1992; accepted 22 April 1993.

1. Kernell, D., Eerbeek, O. & Verhey, B. A. *Exp Brain Res.* **50**, 220-227 (1983).
2. Miledi, R., Parker, I. & Zhu, P. H. *J. Physiol. Lond.* **333**, 655-679 (1982).
3. Armstrong, C. M., Bezanilla, F. M. & Horowitz, P. *Biochim. biophys. Acta* **267**, 605-608 (1972).
4. Adams, B. & Beam, K. *FASEB J.* **4**, 2809-2816 (1990).
5. Rios, E. & Pizarro, G. *Physiol. Rev.* **71**, 849-908 (1991).
6. Catterall, W. A. *Cell* **64**, 871-874 (1991).
7. Kotsias, B. A., Muchnik, S. & Paz, C. A. O. *Am. J. Physiol.* **250**, C40-C46 (1986).
8. Dulhunty, A. F. & Gage, P. W. *J. Physiol. Lond.* **399**, 63-80 (1988).
9. Oz, M. & Frank, G. B. *J. Pharmac. exp. Ther.* **257**, 575-581 (1991).
10. Arreola, J., Calvo, J., Garcia, M. C. & Sanchez, J. A. *J. Physiol., Lond.* **393**, 307-330 (1987).
11. Cheng, H.-C., Van Patten, S. M., Smith, A. J. & Walsh, D. A. *Biochem J.* **231**, 655-671 (1985).
12. Lee, K. S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3941-3945 (1987).
13. Fedida, D., Noble, D. & Spindler, A. J. *J. Physiol., Lond.* **405**, 439-460 (1988).
14. Pietrobon, D. & Hess, P. *Nature* **346**, 651-655 (1990).
15. Artalejo, C. R., Rossie, S., Perlman, R. L. & Fox, A. P. *Nature* **358**, 63-66 (1992).
16. Feldmeyer, D., Meizer, W., Pohl, B. & Zollner, P. *J. Physiol. Lond.* **425**, 347-367 (1990).
17. Garcia, J., Avila-Sakar, A. J. & Stefani, E. *Pflügers Arch.* **416**, 210-212 (1990).
18. Schwartz, L., McCleskey, E. W. & Almers, W. *Nature* **314**, 747-751 (1985).
19. Lawrence, J. C. & Catterall, W. A. *J. biol. Chem.* **256**, 6223-6229 (1981).
20. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch.* **391**, 85-100 (1981).
21. House, C. & Kemp, B. E. *Science* **238**, 1726-1728 (1987).

ACKNOWLEDGEMENTS. We thank M. Binder for discussions of regulation of contractile force by tetanic stimulation. This work was supported by grants from the NIH, the Muscular Dystrophy Association, and Miles Laboratories. A. S. was a Research Fellow of the Heart and Stroke Foundation of Canada.

## Skewed T-cell receptor repertoire in genetically identical twins correlates with multiple sclerosis

Ursula Utz\*, William E. Biddison\*, Henry F. McFarland\*, Dale E. McFarlin\*, Marjorie Flerlage\* & Roland Martin†

\* Neuroimmunology Branch, Bldg 10/Rm 5B-16, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, Maryland 20892, USA

† Department of Neurology, Medical School, University of Tübingen, 7400 Tübingen, Germany

ALTHOUGH the cause of multiple sclerosis (MS) is unknown, it is thought to involve a T cell-mediated autoimmune mechanism. Susceptibility to the disease is influenced by genetic factors such as genes of the HLA and T-cell receptor (TCR) complex<sup>1-6</sup>. Other evidence for a genetic influence includes the low incidence in certain ethnic groups<sup>7</sup>, the increased risk if there are affected family members<sup>8</sup> and the increased concordance rate for disease in monozygotic twin pairs (26%)<sup>9</sup>, compared to dizygotic twins. Epidemiological studies indicate that there may be an additional role for environmental factors. Although the target antigen(s) are not yet identified, several myelin or myelin-associated proteins have been suspected<sup>10-12</sup>, among them myelin basic protein. A lack of genetically comparable controls has impaired the analysis of the T-cell response in MS patients and caused disagreement on TCR usage in the disease<sup>13-15</sup>. Here we analyse the role of TCR genes in MS by comparing TCR usage in discordant versus concordant monozygotic twins in response to self and foreign antigens. We find that after stimulation with myelin basic protein or tetanus

TABLE 1 Summary TCR  $V\alpha$  chain usage in response to MBP stimulation

| Twin sets  | $V\alpha$ chains used by $\geq 60\%$ of T cells |      |             |      |
|------------|-------------------------------------------------|------|-------------|------|
|            | Set 1                                           |      | Set 2       |      |
| Control    | <b>12</b>                                       | (N)  | <b>8</b>    | (N)  |
|            | <b>12, 2</b>                                    | (N)  | <b>8, 6</b> | (N)  |
| Concordant | <b>2, 3</b>                                     | (MS) | <b>8</b>    | (MS) |
|            | <b>2, 8</b>                                     | (MS) | <b>8</b>    | (MS) |
| Discordant | <b>5</b>                                        | (N)  | 3, 10, 8    | (N)  |
|            | <b>8, 2</b>                                     | (MS) | <b>8</b>    | (MS) |

The predominant  $V\alpha$  chains that together make up  $\geq 60\%$  of the TCR  $V\alpha$  chain repertoire in the bulk cultures stimulated with MBP are listed for each individual. The most abundant  $V\alpha$  chain represented is designated in bold. N, normal individuals; MS, affected individuals.

toxoid, control twin sets as well as concordant twin sets select similar  $V\alpha$  chains. Only the discordant twin sets select different TCRs after stimulation with antigens. Thus exogenous factors or the disease shape the TCR repertoire in MS patients, as seen by comparison with unaffected genetically identical individuals. This skewing of the TCR repertoire could contribute to the pathogenesis of MS and other T-cell-mediated diseases.

To eliminate the influence of genetic factors, we investigated six genetically identical, monozygotic twin sets. Two were concordant (both affected), two discordant (one twin affected) for MS. Control twin set (1), was discordant for a psychiatric disease (bipolar disorder); control (2) was clinically healthy.

Bulk cultures of peripheral blood lymphocytes (PBL) from each individual were repeatedly stimulated *in vitro* with myelin basic protein (MBP) or tetanus toxoid and their specificity was confirmed in proliferation assays before TCR analysis. Multiple stimulation of bulk cultures should lead to the selection of predominant responding T cell clonotypes. The resulting shifts in the expressed TCR distribution profiles were determined by polymerase chain reaction (PCR) using  $V\alpha$  chain family-specific primers. Some of the resulting products of amplification were

cloned and sequenced and proved to be members of the selected  $V\alpha$  chain families. Identical  $V\alpha$  and joining-region sequences were detected in MBP-specific T-cell lines from the same individuals (data not shown). A pool of allogeneic cells served as control antigen, showing that broad stimulation of PBL results

in expansion of T cells expressing all  $V\alpha$  chains examined.

No predominant TCR  $V\alpha$  expression was detected before antigen stimulation. Although not all TCR  $V\alpha$  chains were equally abundant, none of them contributed more than 30%. The distribution profiles appeared very similar between two siblings and varied only slightly between different twin sets (Fig. 1a and b). No differences were observed in concordant versus discordant twin sets.  $V\alpha$  2, 3, 7, 8 and 15 were most frequently used, regardless of disease status or HLA haplotypes (legend to Fig. 1). Recent reports confirm similar TCR  $V\beta$  repertoires for identical twins<sup>16,17</sup>.

Repeated stimulation with MBP resulted in the overrepresentation of one to three  $V\alpha$  chains in 11 of the 12 individuals (Fig. 2a-c). Both members of the non-MS control twin sets selected the same  $V\alpha$  families in response to MBP. In control (1) 50–70% of T cells expressed  $V\alpha$  12; in control (2), 70–80% expressed  $V\alpha$  6 and 8 (Fig. 2a). In contrast, individuals of all discordant twin sets selected largely different  $V\alpha$  chains in response to MBP. The affected individuals of both discordant sets expressed predominantly  $V\alpha$  8 (about 40% for (1), and about 70% for (2)), whereas unaffected twin (1) preferentially used  $V\alpha$  5 (>60%) and twin

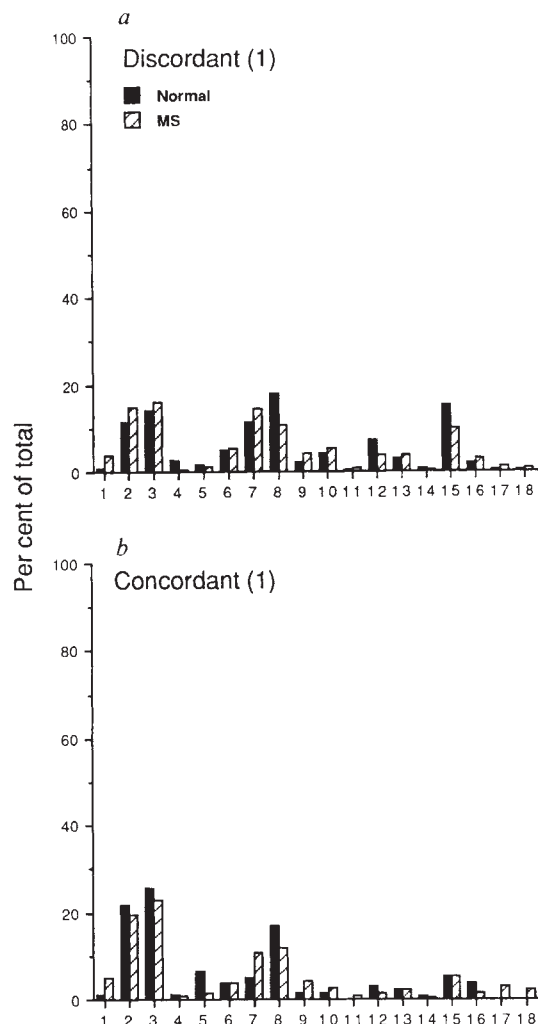


FIG. 1 TCR  $V\alpha$  chain distribution profiles in unstimulated PBL from two monozygous twin sets. a, Discordant twin set (1); b, concordant twin set (1). Lanes 1–18 represent the percentage of each  $V\alpha$  family in comparison with the total number of  $V\alpha$ -bearing T cells.

**METHODS.** Total cellular RNA was prepared from  $5 \times 10^6$  frozen PBL<sup>19</sup> and transcribed into cDNA using oligo(dT) and reverse transcriptase according to the manufacturer's instructions (Pharmacia). cDNA equivalent to 0.5–1.0  $\mu$ g total cellular RNA was used for each individual amplification. 18  $V\alpha$ -chain family-specific oligonucleotides<sup>20</sup> were used as 5' primers ( $C_{\alpha-ext}$  5'-AGCGTCATGAGCAGATTA-3') annealing in the  $\alpha$ -chain constant region as 3' primer. PCR was carried out in 20  $\mu$ l with 0.5 U AmpliTaq, 200  $\mu$ M of each dNTP and 5  $\mu$ M of each primer in manufacturer's buffer (Perkin Elmer Cetus). All reagents except the  $V\alpha$  primers were set up in individual batches for each amplification series. The applied PCR conditions (50 s at 94 °C, 5–20 s at 58 °C, 50 s + 1 s at 72 °C for 30 cycles followed by 10 min at 72 °C) were very stringent and excluded cross-annealing. PCR reactions were separated on 1% agarose gels and blotted onto Nylon membranes (Schleicher & Schuell). Band sizes ranged from 700 to 800 bp. Membranes were hybridized at 42 °C with [ $\gamma$ -<sup>32</sup>P]ATP end-labelled  $C_{\alpha}$ -probe 5'-TTAGAGTCTCTCAGCTGG-3', annealing in the  $\alpha$ -chain constant region upstream from the 3' primer  $C_{\alpha-ext}$ . After consecutive washings with  $3 \times$  SSC,  $10 \times$  Denhardt's, 5% SDS, 10 mM sodium phosphate and  $1 \times$  SSC, 1% SDS at 50 °C, blots were exposed on X-ray film overnight and analysed using a computerized automated scanning system (AMBIS Radioanalytic Imaging System). All signals for the 18  $V\alpha$  families together were scored as 100%. Each PCR analysis was repeated at least once.

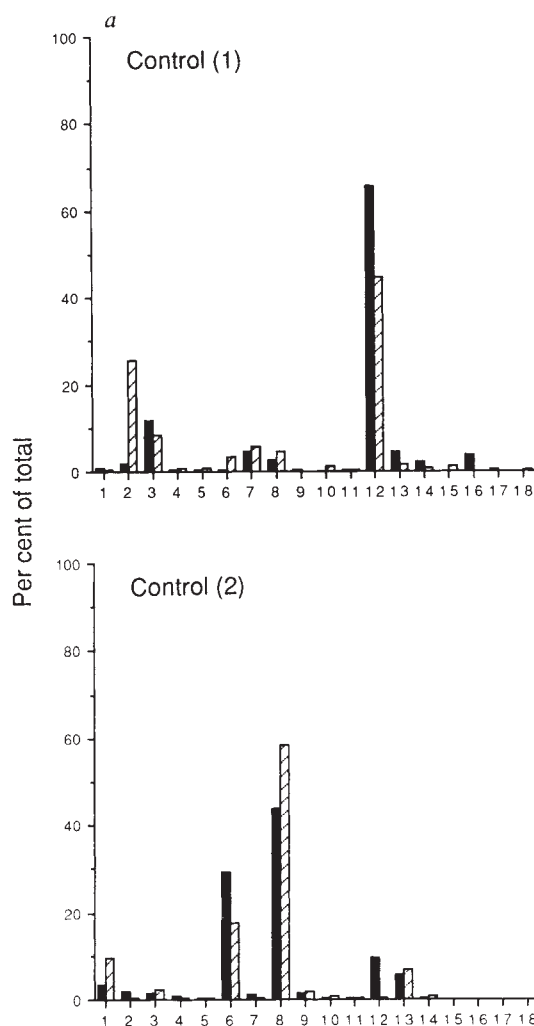


FIG. 2 TCR  $V\alpha$  chain distribution profiles of PBL bulk cultures after stimulation with MBP. a, Profiles of two non-MS control twin sets. (HLA control (1): A1,11; B7,8; Cw7/-; DR15,17; DQ2,6; DRw52/-; HLA control (2): A30,31; B13,60; Cw3/-; DR4,7; DQ2,8; DRw53/-). b, Profiles of two discordant MS twin sets. Affected individuals are marked with MS. (HLA discordant (1): A1,3; B14,56; Cw1/-; DR4,6(13); DQ1(6),3(7); DR52,53; HLA discordant (2): A2/-; B27,60; Cw2,3; DR1,4; DQ5,8; DRw53). c, ►



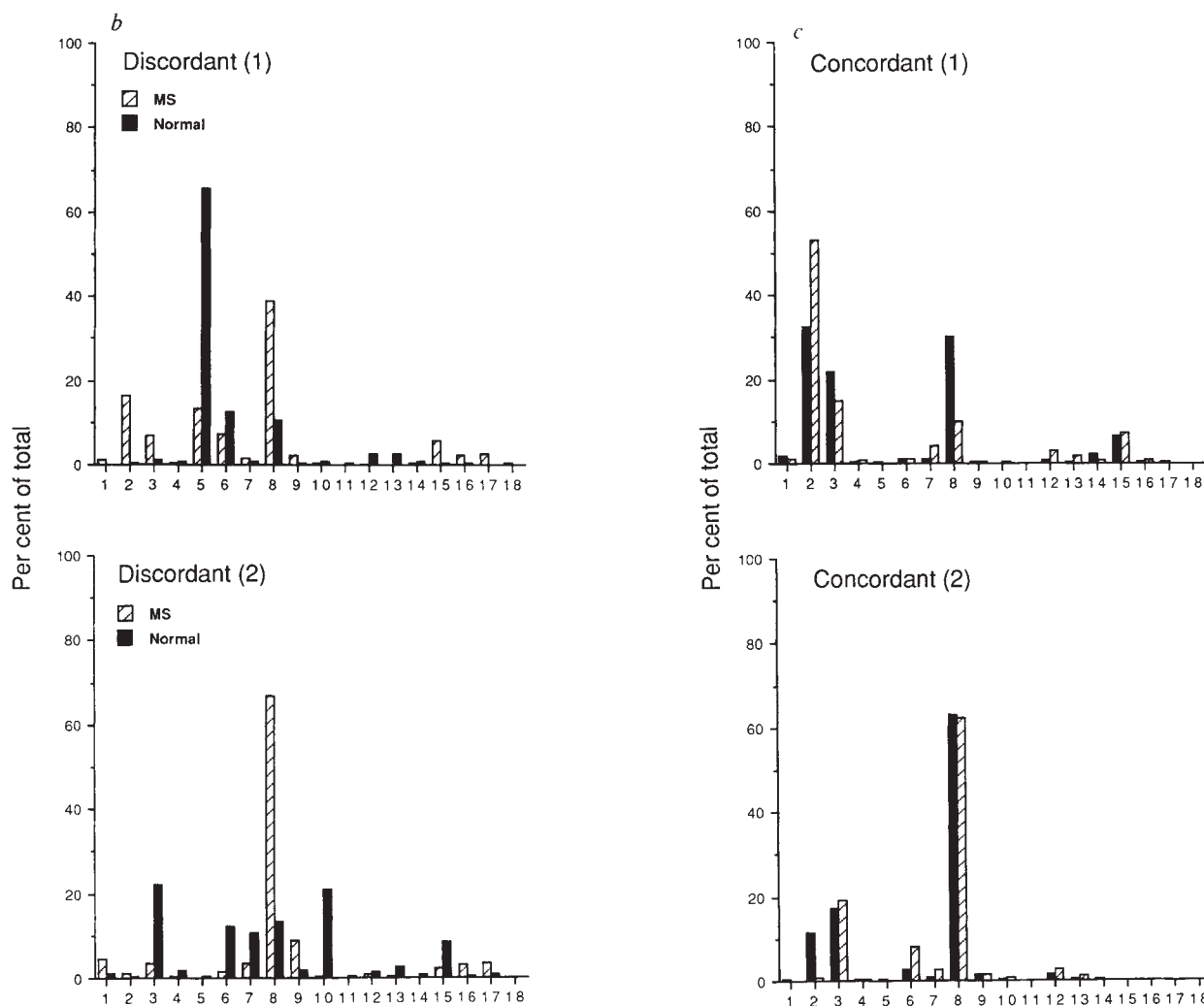
(2) used a combination of *V* $\alpha$ 3, 6, 7, 8, 10 and 15 (Fig. 2*b*). In the concordant twin sets, both twins expressed predominantly *V* $\alpha$ 2, 3 and 8, although in varying amounts, accounting for over 80% of the TCR detected (Fig. 2*c*). A comparison of the *V* $\alpha$  chains that comprise  $\geq 60\%$  of the total TCR expressed in all bulk cultures is shown in Table 1.

To determine whether the selection of divergent TCR *V* $\alpha$  chains by discordant twin pairs is a phenomenon restricted to MBP responses, we analysed *V* $\alpha$  chain profiles after stimulation with tetanus toxoid, a foreign antigen not expected to be involved in the pathogenesis of MS. Again, usually one to four *V* $\alpha$  chains were overrepresented in each individual. Remarkably, stimulation with tetanus toxoid induced again the selection of distinct *V* $\alpha$  chains in the two discordant twin sets. These *V* $\alpha$  chains differed from the ones used in response to MBP (representative results shown for discordant set (1) in Fig. 3*a*). The preferential expression of certain *V* $\alpha$  chains was therefore antigen-specific and did not reflect the outgrowth of the same T-cell populations, independent of antigen stimulation. The unaffected individual of discordant set (1) selected *V* $\alpha$ 8 in response to the toxoid, revealing the presence of potentially responding *V* $\alpha$ 8-

positive T cells. The selection of *V* $\alpha$ 5 in response to MBP, instead of *V* $\alpha$ 8, can therefore not be explained by a lack of these T cells. The twins of concordant set (1) shared *V* $\alpha$  chains in their response to tetanus toxoid (Fig. 3*b*). MBP-specific T-cell lymphocytes from discordant set (2) that had been generated independently produced results similar to those shown. In both instances the affected individual predominantly used *V* $\alpha$ 8, which the unaffected individual did not.

Twins discordant for MS therefore use different TCRs in response both to an autoantigen (MBP) and a foreign antigen (tetanus toxoid), whereas affected concordant twins use the same TCRs in response to each of the antigens. A preferential *V* $\alpha$  usage after stimulation with a particular antigen is not unexpected, especially when immunodominant epitopes are present as in MBP<sup>18</sup>. Multiple stimulations with a pool of allogeneic PBL stimulated a wide variety of different T-cell clones with no apparent differences in *V* $\alpha$  gene expression (Fig. 3*c* and *d*). The differential TCR usage in individuals from discordant twin pairs is therefore not due to a gap in their repertoire.

The shaping of the T-cell repertoire is thought to be influenced by the immunogenetic composition of an individual as well as



Profiles of two concordant MS twin sets. Both individuals are affected with MS. (HLA concordant (1): A2,23; B14,44; Cw- DR2,4; DQ1,3; HLA concordant (2): A24,32; B35,51; Cw4/-; DR1,6; DQ1,2;). PBL from all six twin sets were stimulated weekly in bulk cultures as described<sup>21</sup>. All cultures were tested in proliferation assays for their antigen specificity before this analysis and were found to be MBP-specific with an s.i. > 5. The affected members of the six twin sets had clinically definite MS.

The unaffected twins lacked a history of significant neurological symptoms and had normal neurological examinations. Upon MRI examinations all affected individuals showed changes consistent with MS; MRI were normal in all unaffected members and control twins. Their ages ranged from 35 to 43 years. The duration of disease ranged from 12 to 23 years. Disability in the affected twins ranged from a score of 5.5 to 8.0 on the expanded disability status scale (EDSS)<sup>22</sup>.

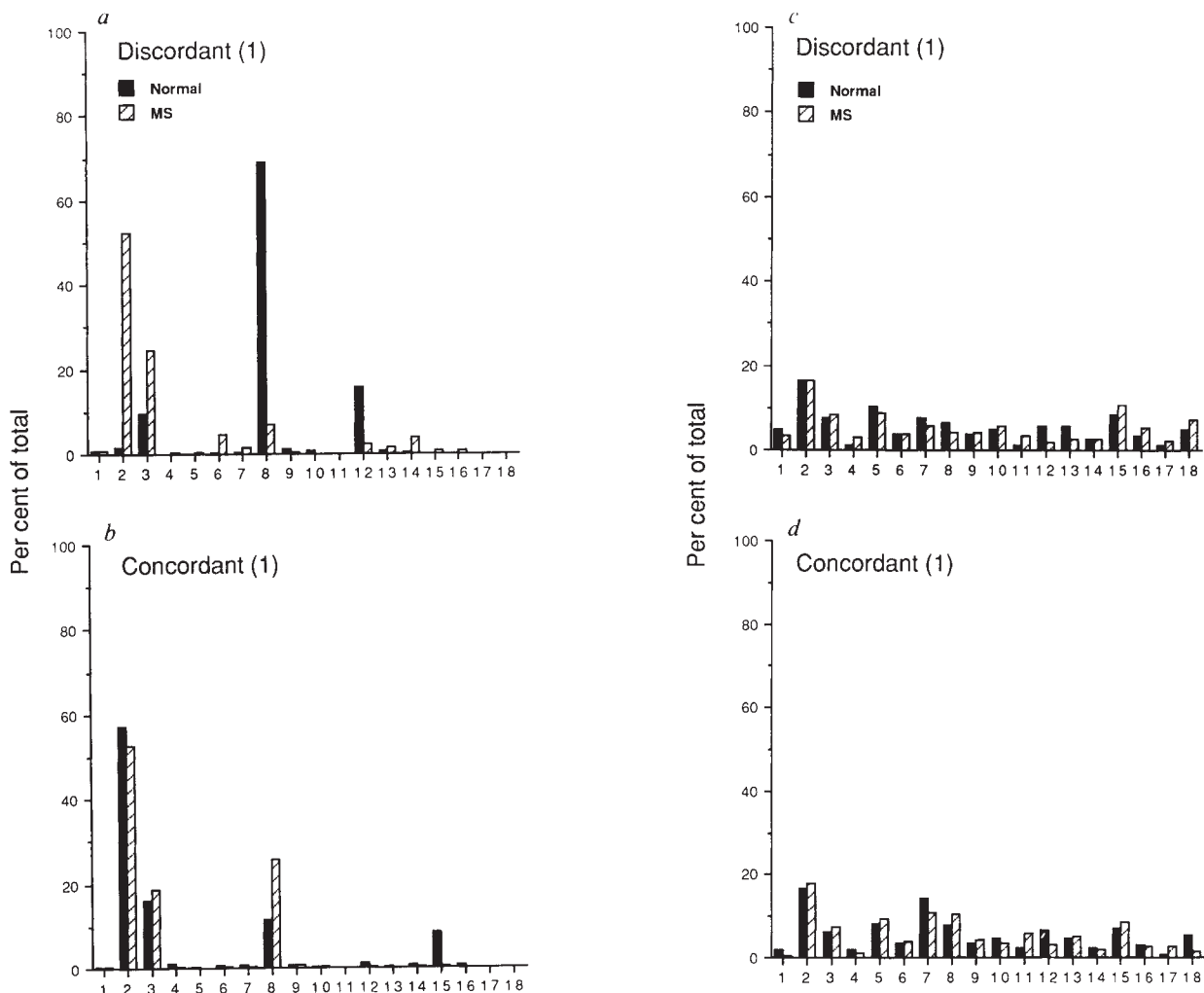


FIG. 3 TCR V $\alpha$  chain distribution profiles of PBL bulk cultures after *in vitro* stimulation with tetanus toxoid. a, Discordant pair (1); b, concordant pair (1). PBL were stimulated eight times with  $10 \mu\text{g ml}^{-1}$  tetanus toxoid presented on autologous PBL. Cultures were tested in proliferation assays beforehand. They were found to be tetanus toxoid-

specific, with s.i. between 2 and 22. c and d, TCR V $\alpha$ -chain distribution profiles of PBL bulk cultures after *in vitro* stimulation with an allogeneic pool of PBL. c, Discordant pair (1); d, concordant pair (1). Cultures previously tested in proliferation assays were specific for the allogeneic PBL pool, with s.i. between 2.6 and 6.

by environmental factors encountered during ontogeny of the immune system. The genetic influences should be similar in monozygotic twins. A discordance for MS in monozygotic twins would be due to environmental factors or somatic events. We cannot determine whether the T-cell repertoires in the twins studied were different before the disease manifested and thus a predisposing factor, or whether the disease process altered them. Our data do indicate, however, that the expressed TCR repertoire differs within discordant twin pairs, but not within concordant or control twin pairs. Those differences do not seem to stem from pre-activated 'bystander' cells which are carried along, because all bulk cultures were antigen-specific and because identical V $\alpha$  and J $\alpha$  sequences were detected in V $\alpha$ 8-positive T cells from the bulk cultures versus MBP-specific TCL (manuscript in preparation). These changes in the repertoire therefore reflect a general skewing that becomes visible not only after stimulation with a suspected target antigen, but also an unrelated foreign antigen like tetanus toxoid. Such antigen nonspecific changes can be caused by superantigens, leading to similar skewing of the TCR repertoire in identical twin pairs if both encounter the antigen, but result in divergence if only encountered by one individual. It is questionable, however, whether any environ-

mental influence can induce a permanent skewing. A chronic T-cell-mediated disease could potentially maintain such an alteration after its initial induction. The observations made for MS twins might therefore generally apply to discordant identical twins suffering from other T-cell-mediated autoimmune diseases such as diabetes, thyroiditis or rheumatoid arthritis. Further investigation of monozygotic twins affected with autoimmune diseases other than MS are needed to determine whether these findings are specific for MS or are a more general phenomenon associated with auto-immunity. □

Received 16 September 1992; accepted 13 April 1993.

1. Vartdal, F., Sollid, L. M., Vandvik, B., Markussen, G. & Thorsby, E. *Hum. Immun.* **25**, 103–110 (1989).
2. Tiwari, J. L. & Terasaki, P. I. *HLA and Disease Associations* 152–167 (Springer, New York, 1985).
3. Cohen, D. et al. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1774–1778 (1984).
4. Olerup, O. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7113–7117 (1989).
5. Beall, S. S. et al. *J. Neuroimmun.* **21**, 59–66 (1989).
6. Seboun, E. et al. *Cell* **57**, 1095–1100 (1989).
7. Waksman, B. H. & Reynolds, W. E. *Proc. Soc. exp. Biol. Med.* **175**, 282–294 (1984).
8. Sadovnick, A. D., Baird, P. A. & Ward, R. H. *Am. J. med. Genet.* **29**, 533–541 (1988).
9. Ebers, G. C. et al. *New Engl. J. Med.* **315**, 1638–1642 (1986).
10. Johnson, D. et al. *J. Neuroimmun.* **13**, 99–108 (1986).
11. Sun, J., Link, H., Olsson, T., Xiao, B.-G. & Andersson, G. *J. Immun.* **146**, 1490–1495 (1991).
12. Allegretta, M., Nicklas, J. A., Sriram, S. & Albertini, R. *J. Science* **247**, 718–721 (1990).



13. Wucherpennig, K. W. *et al. Science* **248**, 1016-1019 (1990).
14. Kotzin, B. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 9161-9164 (1991).
15. Ben-Nun, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 2466-2470 (1991).
16. Gulwani-Alkolkar, B. *et al. J. exp. Med.* **174**, 1139-1146 (1991).
17. Malhotra, U., Spielman, R. & Concannon, P. J. *Immun.* **149**, 1802-1808 (1992).
18. Ota, K. *et al. Nature* **346**, 183-187 (1990).
19. Chirgwin, J. M., Przybala, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5301 (1979).
20. Oksenberg, J. R. *et al. Nature* **345**, 344-346 (1990).
21. Martin, R. *et al. J. Immun.* **148**, 1359-1366 (1992).
22. Kurtzke, J. *Neurology* **33**, 1444-1449 (1983).

ACKNOWLEDGEMENTS. We thank F. Torry for his assistance in obtaining twin pairs; the leukapheresis unit and in particular T. Simonis and the HLA Laboratory, NIH, for their help; and A. Singer and L. A. Matis for critically reading the manuscript. We wish to dedicate this publication to D.E.McF. who died suddenly after completion of this manuscript.

## ATP-dependent glutathione S-conjugate 'export' pump in the vacuolar membrane of plants

Enrico Martinoia, Erwin Grill, Roberto Tommasini, Klaus Kreuz\* & Nikolaus Amrhein

Institute of Plant Sciences, Swiss Federal Institute of Technology, Sonneggstrasse 5, ETH-Z, CH-8092 Zürich, Switzerland  
\* Ciba AG, CH-4002 Basel, Switzerland

PLANTS are exposed to many potentially phytotoxic foreign compounds, such as microbial toxins and agrochemicals (xenobiotics). Detoxification and elimination of these compounds within or from the cell is a prerequisite for their survival. Metabolism and detoxification of xenobiotics are remarkably similar in plants and animals and can generally be divided into three phases<sup>1,2</sup>. In the first phase, a foreign compound may be oxidized, reduced or hydrolysed to introduce or reveal a functional group. In a second step, the activated xenobiotic is conjugated to either glutathione, glucuronate (animals), or malonyl or glucosyl moieties (plants) by the respective transferases. In animals the third step, excretion of conjugated xenobiotics to the extracellular medium, is mediated by a specific ATPase<sup>3,4,5</sup>. In plants, instead of excretion, conjugates of xenobiotics appear to be stored in the large central vacuole<sup>6</sup>, but it is not known how they are transported into this organelle. We show here that glutathione S-conjugate uptake into the vacuole is mediated by a specific ATPase which is remarkably similar to the glutathione S-conjugate export pumps in the canalicular membrane of mammalian liver.

The glutathione S-conjugates of *N*-ethylmaleimide (NEM-GS) and of metolachlor (metolachlor-GS), a glutathione conjugate of a chloroacetanilide herbicide, were barely taken up by isolated barley mesophyll vacuoles in the absence of Mg-ATP. In the presence of Mg-ATP, however, uptake was markedly enhanced (25-80-fold) (Fig. 1). Within the timespan tested, no degradation of the conjugates was detected. Uptake of NEM-GS as well as of metolachlor-GS into vacuoles is a saturable process that follows Michaelis-Menten kinetics. The apparent  $K_m$  values were found to be about 500  $\mu$ M and 40-60  $\mu$ M for NEM-GS and metolachlor-GS, respectively. The maximal transport rates of NEM-GS and metolachlor-GS are similar: 1.6 and 1.4  $\text{nmol} \times 10^{-7}$  vacuoles per minute, respectively.

Various glutathione conjugates, such as metolachlor-GS, biman-GS, simetryn-GS (an *S*-triazine herbicide derivative) and dinitrobenzene-GS inhibited the uptake of NEM-GS (Table 1). Inhibition of NEM-GS uptake both by biman-GS and metolachlor-GS was competitive. The  $K_i$  values were 75  $\mu$ M for metolachlor-GS and 300  $\mu$ M for biman-GS. These results suggest that a single carrier is responsible for the transfer of the different

TABLE 1 Effect of glutathione and conjugates, L-cysteine-NEM and amino acids on NEM-GS uptake

| Treatment               | Per cent of control<br>$\pm$ s.d. | <i>n</i> |
|-------------------------|-----------------------------------|----------|
| Control (+ 5 mM Mg ATP) | 100                               | 3        |
| 2.5 mM glutamic acid    | 105                               | 2        |
| 2.5 mM cysteine         | 96 $\pm$ 5                        | 3        |
| 2.5 mM cysteine-NEM     | 102 $\pm$ 7                       | 3        |
| 1 mM GSH                | 89 $\pm$ 3                        | 3        |
| 1 mM GSSG               | 75 $\pm$ 3                        | 3        |
| 1 mM dinitrobenzene-GS  | 21 $\pm$ 4                        | 3        |
| 1 mM biman-GS           | 43 $\pm$ 6                        | 3        |
| 1 mM simetryn-GS        | 33 $\pm$ 5                        | 3        |
| 1 mM metolachlor-GS     | 20 $\pm$ 1                        | 3        |

Isolated vacuoles were incubated in the presence of NEM-GS (0.4 mM), 5 mM Mg ATP with the compounds listed. Means of two or three experiments (*n*)  $\pm$  s.d., each with five replicates, are shown. Uptake rates were calculated by subtracting the radioactivity taken up after 2 min from the value measured after 20 min. 100% values ranged from 0.73  $\pm$  0.067 to 1.12  $\pm$  0.09  $\text{nmol} \times 10^{-7}$  vacuoles  $\text{min}^{-1}$  for the different experiments. Glutathione S-conjugates were generated by incubating 0.1 M glutathione (reduced; GSH) with 0.12 M of either monobromobimane<sup>22</sup> or chloro-2,4-nitrobenzene in the presence of 0.2 M Tris-HCl, pH 8, 1 mM EDTA for 3 h at 50 °C (pH readjusted to pH 8 with 1 M NaOH). The incubation was repeated after administration of 2-mercaptoethanol (0.05 M) to remove excess reagent. The complete S-derivatization of GSH was confirmed by HPLC analysis with sulphhydryl-specific detection<sup>23</sup>. Simetryn-GS was a gift from Ciba. Derivatives of 2-mercaptoethanol had no effect on the transport of glutathione conjugates.

glutathione conjugates. A weaker inhibition was observed for oxidized glutathione, whereas 1 mM reduced glutathione had only a marginal effect (Table 1). Glutamic acid or cysteine, as well as NEM-cysteine are not inhibitory. By analogy with other vacuolar uptake systems<sup>7</sup>, glutathione conjugate accumulation may be driven by a proton antiport or, owing to the acidic nature of glutathione, be the consequence of the positive membrane potential generated by one of the known proton pumps<sup>8,9</sup>. But bafilomycin, a very potent and specific inhibitor of the tonoplast  $\text{H}^+$ -ATPase<sup>10</sup>, failed to inhibit the ATP-dependent uptake of NEM-GS (Table 2), whereas ATP-dependent proton pumping, as well as the ATP-dependent uptake of malate, which is thought to be a secondary active process<sup>11</sup>, were completely inhibited by the same concentration of bafilomycin. The GS-conjugate transport was also not driven by pyrophosphate, the substrate for the vacuolar  $\text{H}^+$ -pyrophosphatase (data not shown). These results indicate that the energy-dependent uptake of NEM-GS is not linked to either of the known vacuolar proton pumps. The GSSG uptake has properties similar to those of the transport of glutathione conjugates and is furthermore inhibited by these conjugates (results not shown). In contrast, uptake of reduced glutathione is only marginal (15 to 20% of GSSG). NEM-GS uptake is stimulated by 3 mM GTP to about 80% of the maximal uptake rate, and about 50% by 3 mM UTP; the non-hydrolysable ATP analogue AMP-PNP (adenylyl imidodiphosphate) does not stimulate, but rather inhibits the ATP-stimulated uptake (Table 2). Thus hydrolysis of ATP is a prerequisite for the uptake of NEM-GS. Strong inhibition of NEM-GS uptake by vanadate ( $\text{IC}_{50}$  60  $\mu$ M) also argues against an involvement of the tonoplast  $\text{H}^+$ -ATPase. Vacuolar  $\text{H}^+$ -ATPases are not inhibited by vanadate<sup>8,9</sup>, although this anion inhibits the plasma membrane  $\text{H}^+$ - and  $\text{Ca}^{2+}$ -ATPases<sup>12</sup> which form phosphorylated intermediates during ATP hydrolysis. A further distinction of the glutathione-conjugate uptake system from either  $\text{H}^+$ -pumping in intact vacuoles<sup>13</sup> or from secondary activated processes<sup>14</sup> is its higher affinity for Mg ATP ( $K_m$  for Mg ATP is about 100  $\mu$ M). These observations suggest that glutathione

TABLE 2 Effect of inhibitors and of adenylyl imidodiphosphate on NEM-GS uptake,  $H^+$  pumping and malate transport into barley mesophyll vacuoles

| Treatment                         | NEM-GS uptake         |                      | <i>n</i> | $H^+$ pumping<br>(% of maximal activity) | Malate uptake         |                      | <i>n</i> |
|-----------------------------------|-----------------------|----------------------|----------|------------------------------------------|-----------------------|----------------------|----------|
|                                   | nmol $\times 10^{-7}$ | vacuoles min $^{-1}$ |          |                                          | nmol $\times 10^{-7}$ | vacuoles min $^{-1}$ |          |
| Control(–Mg-ATP)                  | 0.03 $\pm$ 0.012      | (4 $\pm$ 3)          | 6        | 0                                        | 6.6 $\pm$ 0.4         | (23)                 | 2        |
| +Mg-ATP                           | 0.84 $\pm$ 0.06       | (100)                | 6        | 100                                      | 25.3 $\pm$ 2.6        | (100)                | 2        |
| +Mg-AMP-PNP                       | 0.01 $\pm$ 0.01       | (1)                  | 2        | ND                                       | ND                    |                      |          |
| +Mg-ATP + 0.1 $\mu$ M bafilomycin | 0.85 $\pm$ 0.08       | (102 $\pm$ 6)        | 4        | 0                                        | 6.5 $\pm$ 0.6         | (24)                 | 2        |
| +Mg-ATP + 1 mM vanadate           | 0.075 $\pm$ 0.015     | (8 $\pm$ 4)          | 4        | 98                                       | 24.8 $\pm$ 2.6        | (98)                 | 2        |
| 0.5 mM Mg-ATP                     | 0.81 $\pm$ 0.08       | (94 $\pm$ 7)         | 4        | ND                                       | ND                    |                      |          |
| 0.5 mM Mg-ATP + 1 mM AMP-PNP      | 0.21 $\pm$ 0.03       | (27)                 | 2        | ND                                       | ND                    |                      |          |

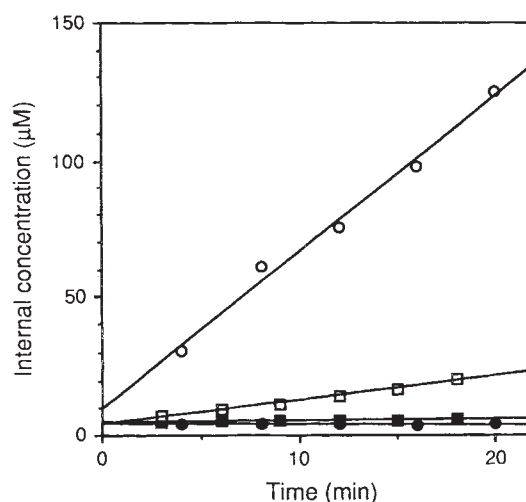
Uptake of NEM-GS (0.4 mM NEM-GS, 3 mM Mg-ATP, unless stated otherwise) and malate (4 mM malate, 5 mM Mg-ATP) was determined as described in the legends of Fig. 1 and Table 1. Absolute values are given for a typical experiment with 5 replicates; values in parentheses represent per cent of maximal activity  $\pm$  s.d. from *n* experiments. Acidification of the vacuolar sap was measured by incubating  $1-3 \times 10^5$  vacuoles in basal medium (Fig. 1 legend) containing 50 mM KCl, 5 mM Mg-ATP, 30  $\mu$ M acridine orange and solutes as indicated. Time-dependent fluorescence quenching was observed at 490 nm. Means of two experiments are presented. ND, not determined.

conjugate uptake into plant vacuoles is mediated by a specific ATPase. Activation of the carrier by a membrane-bound protein kinase could be an alternative. In this case, however, transport of the GS conjugates against a concentration gradient would require the activity of one of the vacuolar proton pumps. Our observation that metolachlor-GS is accumulated even in the presence of bafilomycin (Table 3) supports the view that glutathione conjugates are transferred to the vacuole by a specific ATPase. It can be excluded that the positive membrane potential often observed in isolated vacuoles is the driving force for the uptake of the glutathione conjugates. The membrane potentials in isolated vacuoles are in the range of 5–20 mV (ref. 15). Such a membrane potential would drive, at most, a twofold accumulation. A 5–6-fold accumulation, as observed after one hour (Table 3), would require a membrane potential of more than 40 mV, which can only be reached with an active proton pump. An ATP-stimulated proton/glutathione conjugate exchange mechanism can also be excluded, because 5 mM  $NH_4Cl$  (which dissipates the pH gradient) did not reduce GS-conjugate uptake significantly (93  $\pm$  6% of the control).

The vacuolar ATPase mediating the export of glutathione con-

jugates from the cytosol into the vacuole shows a striking resemblance to the ATPase in the canalicular membrane of liver, which exports glutathione conjugates and oxidized glutathione (GSSG) into the extracellular medium<sup>1,3–5</sup>. In both cases, the glutathione *S*-conjugates of structurally unrelated xenobiotics as well as GSSG are recognized. Furthermore, both systems are inhibited by vanadate, they are not activated by AMP-PNP, and GTP is able partially to replace ATP. The vacuolar membrane appears therefore to have not only a similar function to the canalicular membrane of liver in excreting waste and toxic products, but it has functionally similar transport systems. It is remarkable that the liver cell excretes glutathione conjugates to the extracellular medium for ultimate elimination from the organism, whereas plants sequester glutathione conjugates by compartmentation in a large organelle with very limited metabolic activity. This process can be described as 'storage excretion'<sup>16</sup>. The ATP-dependent glutathione *S*-conjugate export pump from liver has not so far been purified, but it is distinct from two other ATPases mediating the export of toxic or waste compounds in mammalian liver: the bile salt ATPase<sup>17,18</sup> and the multidrug-resistance protein (MDR)<sup>19</sup>. The MDR protein belongs to an ATPase family

FIG. 1 Time-dependent uptake of [ $^{14}C$ ]metolachlor-GS (●, ○) and [ $^{14}C$ ]NEM-GS (■, □) (each 50  $\mu$ M) into isolated barley mesophyll vacuoles incubated in the absence (●, ■) or presence (○, □) of 3 mM Mg-ATP. Means of two determinations. Barley (*Hordeum vulgare* L. cv Gerbel) was grown in a growth cabinet with 12 h fluorescent light (45  $\mu$ mol m $^{-2}$  s $^{-1}$ ) at 22 °C and 12 h dark at 18 °C; relative humidity was 75%. Protoplasts and vacuoles were isolated as described<sup>14,24</sup>. Contamination of the vacuolar preparation with other cell constituents was less than 1% as measured by marker enzyme activities<sup>14</sup>. Uptake of [ $^{14}C$ ]NEM-GS and [ $^{14}C$ ]metolachlor-GS was measured by a slight modification of the method described in refs 11, 14. For each condition and time point, polyethylene microcentrifugation tubes (400  $\mu$ l, capacity) were prepared as follows: 70  $\mu$ l of basal medium (0.4 M sorbitol, 30 mM potassium gluconate, 25 mM HEPES-KOH, pH 7.2) containing 40% (v/v) Percoll, 0.12% bovine serum albumin, and either 2 kBq [ $^{14}C$ ]NEM-GS or 0.5 or 1 kBq [ $^{14}C$ ]metolachlor-GS, as well as 2 kBq  $^3H_2O$ , were put in the bottom of the tube. The concentration of ATP was kept constant by addition of phosphocreatine (10 mM) and creatine phosphokinase (16 units ml $^{-1}$ ).  $Mg^{2+}$  was present at a concentration of 3.3 mM and was added as  $MgSO_4$ . Glutathione conjugate contents at zero time reflect glutathione conjugate of the suspending medium that contaminates the vacuoles after their separation from the medium. Uptake was started by addition of 30  $\mu$ l vacuole suspension. Samples were rapidly overlaid with 200  $\mu$ l silicone oil AR 200 (Fluka, Buchs Switzerland) and 60  $\mu$ l water. The incubation was terminated by centrifugation at 10,000 g for 15 s.  $^3H_2O$  equilibrates rapidly between the medium and the vacuolar space and can be used to quantify the number of vacuoles.  $10^7$  vacuoles correspond to a volume of 160  $\mu$ l and approximately 1 mg chlorophyll in barley mesophyll cells<sup>14</sup>. [ $^{14}C$ ]metolachlor-GS (specific activity 0.203 GBq mmol $^{-1}$ ) and unlabeled metolachlor-GS were synthesized as described<sup>25</sup>.  $^{14}C$ -labelled N-ethylmaleimide-glutathione (NEM-GS) was synthesized by mixing 1.85 MBq  $^{14}C$ -labelled N-ethylmaleimide (1.4 GBq mmol $^{-1}$ ; NEN DuPont, USA) dissolved in 200  $\mu$ l 20 mM Tris-HCl, pH 7.2, with 1.4  $\mu$ mol glutathione dissolved in 200  $\mu$ l of the same buffer. After 1 h at room temperature, 0.5  $\mu$ mol DTT was added. Analysis by TLC showed that the reaction was quantitative. DTT and conjugates of DTT had no effect on the transport of glutathione conjugates.



labelled metolachlor-GS were synthesized as described<sup>25</sup>.  $^{14}C$ -labelled N-ethylmaleimide-glutathione (NEM-GS) was synthesized by mixing 1.85 MBq  $^{14}C$ -labelled N-ethylmaleimide (1.4 GBq mmol $^{-1}$ ; NEN DuPont, USA) dissolved in 200  $\mu$ l 20 mM Tris-HCl, pH 7.2, with 1.4  $\mu$ mol glutathione dissolved in 200  $\mu$ l of the same buffer. After 1 h at room temperature, 0.5  $\mu$ mol DTT was added. Analysis by TLC showed that the reaction was quantitative. DTT and conjugates of DTT had no effect on the transport of glutathione conjugates.



TABLE 3 Accumulation of [ $^{14}\text{C}$ ]metolachlor-GS in isolated barley vacuoles

| Treatment                      | Internal concentration     |                            |
|--------------------------------|----------------------------|----------------------------|
|                                | Experiment 1               | Experiment 2               |
| -Bafilomycin                   | 275 $\pm$ 25 $\mu\text{M}$ | 303 $\pm$ 64 $\mu\text{M}$ |
| +0.1 $\mu\text{M}$ bafilomycin | 258 $\pm$ 53 $\mu\text{M}$ | 263 $\pm$ 22 $\mu\text{M}$ |

Vacuoles were incubated for 60 min in the presence of 50  $\mu\text{M}$  [ $^{14}\text{C}$ ]metolachlor-GS in the presence or absence of bafilomycin. The internal concentration of metolachlor-GS was calculated from the ratio of  $^3\text{H}_2\text{O}$  and [ $^{14}\text{C}$ ]metolachlor in the incubation medium and in the vacuoles. For each experiment, means of five determinations  $\pm$  s.d. are shown. For further experimental details, see legend to Fig. 1.

that includes the periplasmic binding proteins of prokaryotes and bacterial exporters, as well as eukaryotic proteins such as the *STE6* gene product, which mediates the export of a yeast mating pheromone, the CFTR product of the gene that is faulty in cystic fibrosis<sup>20</sup>, and the protein responsible for cadmium tolerance in *Schizosaccharomyces pombe*<sup>21</sup>. It may be that other MDR-like ATPases exist that are responsible for detoxification processes in animals and plants.  $\square$

## Protein kinase C $\alpha$ activates RAF-1 by direct phosphorylation

Walter Kolch<sup>\*,†</sup>, Gisela Heidecker<sup>‡</sup>,  
Georg Kochs<sup>\*§</sup>, Richard Hummel<sup>\*§</sup>,  
Haleh Vahidi<sup>‡</sup>, Harald Mischak<sup>||</sup>,  
Günter Finkenzeller<sup>§</sup>, Dieter Marmé<sup>\*§</sup>  
& Ulf R. Rapp<sup>‡</sup>

\* Biological Research, Goedecke AG, Mooswaldallee 1–9, D-7800 Freiburg, Germany

† Laboratory of Viral Carcinogenesis, NIH/NCI, FCRDC, Bldg 560, Frederick, Maryland 21702, USA

§ Institute of Molecular Cell Biology, University of Freiburg, Goedecke AG, Mooswaldallee 1–9, D-7800 Freiburg, Germany

|| Laboratory of Genetics, NIH/NCI, Bldg 37, Bethesda, Maryland 20892, USA

THE kinase Raf-1 can be activated by treatment of cells with mitogens and by the protein kinase C (PKC) activator 12-*O*-tetradecanoyl-phorbol-13-acetate (TPA) (reviewed in refs 1, 2). Activated Raf-1 triggers a protein kinase cascade by direct phosphorylation of MAP kinase kinase<sup>3–5</sup>, resulting in phosphorylation of ternary complex factor<sup>6</sup> and Jun<sup>7,8</sup> by MAP kinase. Here we investigate the molecular mechanism and biological consequences of PKC $\alpha$ -mediated Raf-1 activation in NIH3T3 fibroblasts. PKC $\alpha$  directly phosphorylates and activates Raf-1 both *in vitro* and *in vivo*. PKC $\alpha$  induces Raf-1 phosphorylation at several sites, including a serine residue at position 499. Mutation of serine at this position or at residue 259 does not abrogate Raf-1 stimulation by a combination of Ras plus the *src* tyrosine kinase Lck, but severely impedes Raf-1 activation by PKC $\alpha$ . Consistent with such a direct interaction is the observation that Raf-1 and PKC $\alpha$  cooperate in the transformation of NIH3T3 cells. The Ser499 phosphorylation site is necessary for this synergism.

Raf-1 was immunoprecipitated from orthophosphate-labelled NIH3T3 cells and PKC $\alpha$  transfectants, which overexpress the

Received 17 December 1992; accepted 14 April 1993.

1. Ishikawa, T. *Trends biochem. Sci.* **17**, 463–468 (1992).
2. Sandermann, H. *Trends biochem. Sci.* **17**, 82–84 (1992).
3. Kobayashi, K., Sogame, Y., Hara, H. & Hazashi, K. *J. biol. Chem.* **265**, 7737–7741 (1990).
4. Kitamura, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 3557–3561 (1990).
5. Akerboom, T. P., Narayanaswami, V., Kunst, M. & Sies, H. *J. biol. Chem.* **266**, 13147–13152 (1991).
6. Schmitt, R. & Sandermann, H. *Z. Naturforsch.* **37c**, 772–777 (1982).
7. Martinoia, E. *Bot. Acta* **105**, 232–245 (1992).
8. Sze, H. A. *Rev. Pl. Physiol.* **36**, 175–208 (1985).
9. Rea, P. A. & Sanders, D. *Physiol. Plant.* **71**, 131–141 (1987).
10. Bowman, E. J., Siebers, A. & Altendorf, K. H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7972–7976 (1988).
11. Martinoia, E., Flüggé, U. I., Kaiser, G., Heber, U. & Heidt, H. W. *Biochim. biophys. Acta* **806**, 311–319 (1985).
12. Serrano, R. A. *Rev. Pl. Physiol. Plant molec. Biol.* **40**, 61–94 (1989).
13. Hedrich, R. & Schroeder, J. A. *Rev. Pl. Physiol. Plant molec. Biol.* **40**, 539–569 (1989).
14. Rentsch, D. & Martinoia, E. *Planta* **184**, 532–537 (1991).
15. Barbier-Brygoo, H. *et al. Biochim. biophys. Acta* **819**, 215–224 (1985).
16. Luckner, in *Secondary Metabolism in Microorganisms, Plants and Animals* 3rd edn 452 (Springer, Berlin, Heidelberg and New York, 1990).
17. Nishida, T., Gatmaitan, Z., Che, M. & Arias, I. M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6590–6594 (1991).
18. Stieger, B., O'Neill, B. & Meier, P. J. *Biochem. J.* **284**, 67–74 (1992).
19. Endicott, J. A. & Ling, V. A. *Rev. Biochem.* **58**, 137–171 (1989).
20. Hyde, S. C. *et al. Nature* **346**, 362–365 (1990).
21. Ortiz, D. F. *et al. EMBO J.* **10**, 3491–3499 (1992).
22. Fahey, R. C., Newton, G. L., Dorian, R. & Kosower, E. M. *Analyt. Biochem.* **107**, 1–10 (1980).
23. Grill, E., Winnacker, E. L. & Zenk, M. H. *Proc. natn. Acad. Sci. U.S.A.* **84**, 439–443 (1987).
24. Kaiser, G., Martinoia, E. & Wiemken, A. Z. *Pflanzenphysiologie* **107**, 103–113 (1982).
25. Fuerst, E. P. & Gronwald, J. W. *Weed Sci.* **34**, 354–361 (1986).

ACKNOWLEDGEMENTS. We thank E. Vogt for technical assistance, and the Swiss National Science Foundation for financial support.

enzyme fourfold (Fig. 1A). In starved NIH cells, Raf-1 and a coprecipitating protein of *M<sub>r</sub>* 120K (ref. 9) showed a low basal phosphorylation level. In contrast, Raf-1 and p120 were hyperphosphorylated, but not shifted, in immunoprecipitates from starved PKC $\alpha$  transfectants. TPA stimulation increased phosphorylation of Raf-1 and p120 in NIH3T3, but not in PKC $\alpha$  cells. But Raf-1 was shifted in both cell lines, suggesting that the Raf-1-shift-inducing phosphorylation occurs at specific sites. Direct phosphorylation of Raf-1 by PKC $\alpha$  *in vitro* was tested with full-length Raf-1 produced in Sf9 cells and an *Escherichia coli* expressed 30K Raf-1 kinase domain peptide, c-raf30K (Fig. 2b). PKC $\alpha$  phosphorylated c-raf30K better than the  $\beta$  or  $\gamma$  isozyme (Fig. 1B). Similar results were obtained with full-length Raf-1 (data not shown).

To investigate phosphorylation *in vivo*, phosphorylation sites were identified for Raf-1 immunoprecipitated from serum-starved or TPA-treated EC0 cells<sup>10</sup>, a Raf-1 overexpressing NIH3T3 clone, which was metabolically labelled with orthophosphate. Phosphopeptide maps of trypsin-digested Raf-1 were compared with the pattern from *in vitro* phosphorylated c-raf30K (Fig. 1C). Stimulation with TPA induced hyperphosphorylation of pre-existing as well as new sites. Several peptides phosphorylated by PKC $\alpha$  *in vitro* were also induced by TPA *in vivo*. These results indicate that Raf-1 is a direct substrate for PKC $\alpha$  *in vitro* as well as *in vivo*.

Raf-1 contains several potential PKC substrate sites<sup>2,9</sup>, including Ser43, Ser259 and a degenerate site at Ser499. Serine at position 499 corresponds to a major regulatory site in Src<sup>11</sup>, namely Tyr416, and therefore was examined first. Synthetic peptide VKS494RWS497GS499QQVEQPT was phosphorylated by PKC $\alpha$ , so were peptides in which Ser494 or Ser497 were replaced by alanine. The Ala499 peptide, however, was not phosphorylated (Fig. 2a).

To verify that Ser499 is a PKC $\alpha$  phosphorylation site, the residue was changed to alanine by site-directed mutagenesis. The mutation, termed S499A, was placed in the context of a full-length Raf-1, the truncated Raf-1 versions BXB<sup>10</sup> and c-raf30K (Fig. 2b). To achieve high expression for *in vivo* <sup>32</sup>P-labelling, BXB and BXBBS499A were transiently transfected into A293 human kidney embryo fibroblasts, which have PKC $\alpha$  as the predominant isozyme (W.K., unpublished data). Immunopurified BXB and BXBBS499A were digested with trypsin plus V8 protease to obtain finer resolution on two-dimensional peptide

† Present address: Institute for Clinical Molecular Biology and Tumor Genetics, GSF, Haematologikum, Marchioninistrasse 25, D-8000 Muenchen 70, Germany.

maps and to allow comparison with the *in vitro* phosphorylated VKS494R/WS497GS499QQVE/QPT peptide (slashes indicate trypsin and V8 cleavage sites). BXB phosphorylation sites represented a subset of Raf-1 sites (data not shown). BXBS499A had lost a TPA-induced phosphorylation site, which was also absent in c-raf30K-S499A phosphorylated by PKC $\alpha$  *in vitro* (Fig. 2c). This peptide co-migrated with the /WS497GS499QQVE/ synthetic peptide (data not shown).

We used the Sf9/baculovirus system to test RafS499A for activation by PKC $\alpha$  in the absence of endogenous wildtype Raf-1 and assayed for autophosphorylation or activity on exogenous substrate. Raf-1 purified from TPA-stimulated Sf9 cells showed a slight increase in kinase activity towards a synthetic peptide substrate. Coexpression of PKC $\alpha$  enhanced the phospho-

transferase activity of Raf-1, which was further elevated when cells were treated with TPA. Immunopurified Raf-1 pre-phosphorylated with PKC $\alpha$  *in vitro* also exhibited increased kinase activity. In contrast, RafS499A was largely refractory to stimulation by PKC $\alpha$  *in vivo* as well as *in vitro* (Fig. 3a).

In the autophosphorylation assay, in addition to mutant RafS499A, RafS259D but not RafS43A was defective for activation by PKC $\alpha$ . In the absence of PKC $\alpha$ , TPA activated wild-type Raf-1 and RafS43A 2- to 3-fold (data not shown). All mutants, however, were stimulated more than tenfold when coexpressed with Ras and the *src* family tyrosine kinase Lck (Fig. 3b). Lck participates in Raf-1 activation in T cells<sup>12</sup> and the combination of Src and Ras accomplished optimal Raf-1 activation in Sf-9 cells<sup>13</sup>.

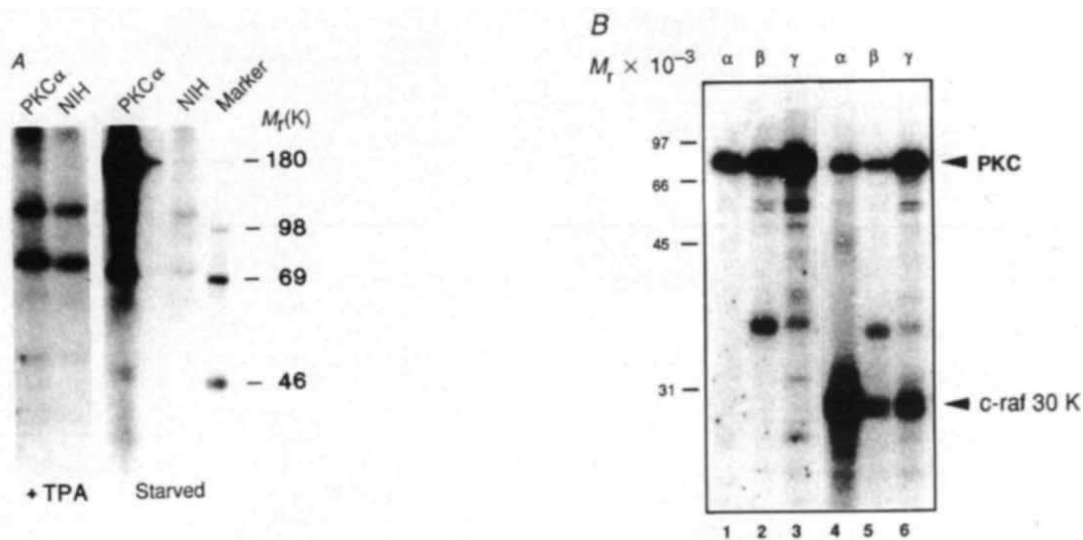


FIG. 1 Raf-1 is a PKC substrate *in vivo* and *in vitro*. A, Phosphorylation at specific sites is required to induce a Raf-1 gel shift. Serum-starved PKC $\alpha$ -transfected (PKC $\alpha$ ) or NIH3T3 control cells (NIH) were metabolically labelled with 1 mCi ml<sup>-1</sup> orthophosphoric acid for 3 h. After 20 min of TPA stimulation (100 ng ml<sup>-1</sup>), Raf-1 was immunoprecipitated from lysates adjusted to be equal in protein content. An unidentified 170K protein coprecipitates with Raf-1 in some cell lines<sup>9</sup>. B, Purified PKCs  $\alpha$ ,  $\beta$  and  $\gamma$  phosphorylate the kinase-negative c-raf30K peptide *in vitro*; c-raf30K is a purified *E. coli* expression protein that includes amino acids 387–595 of Raf-1 (Fig. 2b). Phosphorylations were done in kinase buffer (25 mM Tris-HCl, pH 7.5, 1.32 mM CaCl<sub>2</sub>, 5 mM MgCl<sub>2</sub>, 1 mM EDTA, 1.25 mM EGTA, 1 mM DTT) supplemented with diacylglycerol (1  $\mu$ g ml<sup>-1</sup>), phosphatidylserine (5  $\mu$ g ml<sup>-1</sup>), 10  $\mu$ M ATP and 0.5  $\mu$ Ci [ $\gamma$ -<sup>32</sup>P]ATP. To compare specific activities of PKC preparations, autophosphorylation was tested in the absence of exogenous substrate (lanes 1–3). In lanes 4–6, c-raf30K was added as substrate. C, PKC $\alpha$  phosphorylates the same sites *in vitro* and *in vivo*. Raf-1 was immunoprecipitated from serum-starved [<sup>32</sup>P]orthophosphate-labelled ECO cells which had been stimulated with 100 ng ml<sup>-1</sup> TPA for 30 min or left untreated. The c-raf30K protein was phosphorylated with PKC $\alpha$  *in vitro*. Proteins were separated by SDS-PAGE and blotted onto nitrocellulose membranes. Raf bands were cut out and processed for digestion with trypsin and two-dimensional peptide mapping as described<sup>16</sup>. In the first dimension, peptides were separated by electrophoresis in 1% ammonium carbonate buffer, pH 8.9, for 350 volt-hours. Anode is on the left; origin is at the bottom left-hand corner. The second dimension was ascending chromatography in water: acetic acid: *n*-butanol: pyridine (60:15:75:50) buffer.

METHODS. An EcoRI-XbaI fragment of a human PKC $\alpha$  cDNA clone (provided by J. Knopf) was cloned into the Elneo expression vector<sup>10</sup> and transfected into NIH3T3 cells. Overexpression of PKC $\alpha$  in G418-resistant cell pools was 3–4-fold, as monitored by densitometry of western blots. Human PKC $\alpha$  and  $\beta$ , and rat PKC $\gamma$  were expressed in the baculovirus/Sf9 cell system<sup>17</sup> and purified to homogeneity<sup>18</sup>. To express

the c-raf30K protein, a *Stu*I fragment (nucleotides 1,295–1,917) of the human Raf-1 cDNA was cloned in-frame into the pQE *E. coli* expression plasmid (Qiagen, Germany). Purification of c-raf30K protein from inclusion bodies, the preparation of cell lysates, immunoprecipitation and western blotting were all done as described<sup>9</sup>.



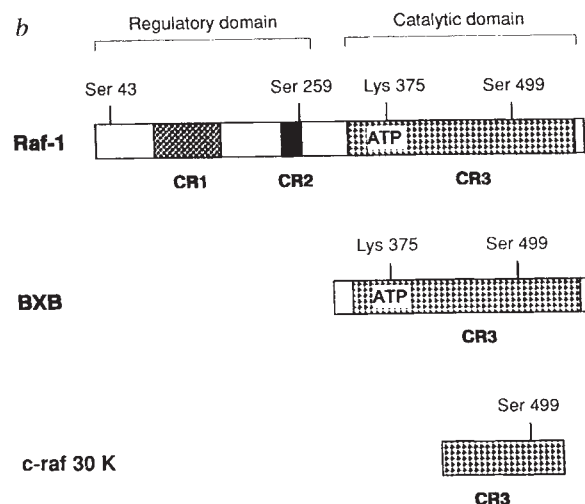
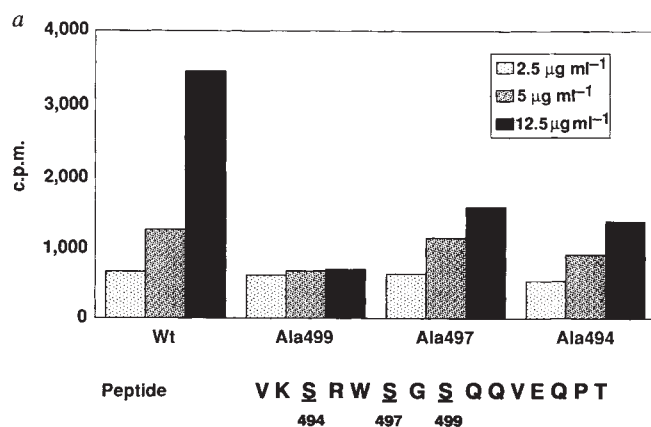


FIG. 2 PKC $\alpha$  phosphorylates Raf-1 on Ser499. **a**, Synthetic peptides (Cambridge Biochem.) were tested as substrates for PKC $\alpha$ . Phosphorylations were done as described for Fig. 1B, except that diacylglycerol and phosphatidylserine were at  $5 \mu\text{g ml}^{-1}$  and  $25 \mu\text{g ml}^{-1}$  respectively. Reactions were incubated for 5 min at  $30^\circ\text{C}$ , stopped with 5% phosphoric acid and filtered through Whatman P81 phosphocellulose filters. After extensive washing with phosphoric acid, phosphorylation of filter-bound peptides was quantitated by Cerenkov counting. The assay was repeated three times with similar results. Replacement of Ser499, but not of adjacent serines at positions 494 and 497, abolished phosphorylation by PKC $\alpha$ . **b**, Schematic representation of Raf-1 proteins used for determining phosphorylation sites and kinase activities. Amino acids, which were changed by site-directed mutagenesis, are indicated. **c**, The Ser499 to Ala point mutation eliminates a TPA-induced Raf-1 phosphorylation site. BXB and BXBS499A cDNAs cloned into the mammalian expression vector pMNC under control of the cytomegalovirus promoter<sup>10</sup> were transiently expressed in A293 cells. Raf proteins were immunoprecipitated from serum-starved or TPA-treated (for 30 min at  $100 \text{ ng ml}^{-1}$ ) cells metabolically labelled with  $1 \text{ mCi ml}^{-1}$  orthophosphoric acid. c-raf30K and c-raf30K-S499A were phosphorylated with PKC $\alpha$  *in vitro*. Phosphopeptides were mapped as in Fig. 1, except that proteins were double-digested with trypsin and V8 protease. Additional phosphopeptides in the c-raf30K maps are probably due to spurious phosphorylation of the partially denatured *E. coli* expression protein.

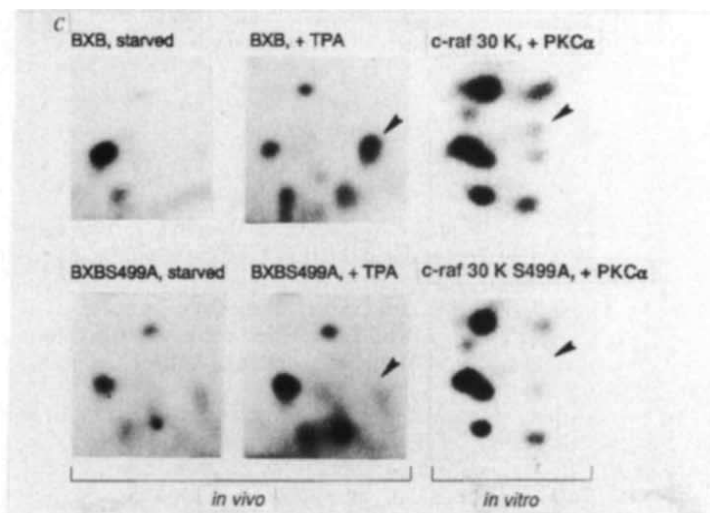
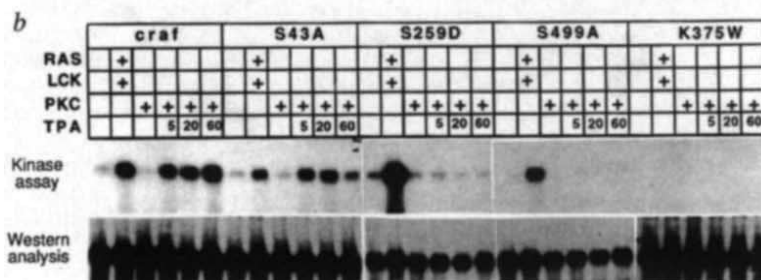
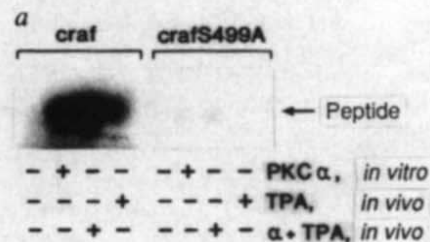


FIG. 3 Mutation of Ser499 or Ser259 specifically impairs activation of Raf-1 by PKC $\alpha$ . Wild-type Raf-1 and mutants RafS43A, RafS259D, RafS499A and RafK375W (Fig. 2b) were cloned into baculovirus vectors (pVL941) and expressed in Sf9 cells<sup>17</sup>. **a**, Raf-1 or RafS499A were expressed alone or in combination with human PKC $\alpha$ . Sf9 cells were treated where indicated with  $100 \text{ ng ml}^{-1}$  TPA for 30 min or otherwise Raf immunoprecipitates were pre-incubated *in vitro* with PKC $\alpha$  in kinase buffer supplemented with  $100 \mu\text{M}$  ATP for 30 min. After extensive washing of immunoprecipitates, Raf-1 kinase activity was measured with a synthetic peptide substrate (CQQFGFQRASDDG) as described<sup>19</sup>. Phosphorylated peptides were separated on 20% SDS-gels and visualized by autoradiography. **b**, Sf9 cells were infected with viruses carrying Raf-1 or RafS43A, RafS259D, RafS499A and RafK375W mutants. Viruses expressing Lck, Ras or PKC $\alpha$  were added as indicated. Before collection, cells were treated with  $200 \text{ ng ml}^{-1}$  TPA for 5, 20 or 60 min. Cells were lysed in RIPA buffer (20 mM Tris-HCl, pH 7.5, 140 mM NaCl, 0.5% deoxycholate, 0.1% SDS, 1% Triton X100, 10% glycerol, 0.5 mM PMSF, 0.01% aprotinin, 0.01% leupeptin) and standardized for protein content. Raf-1 immunoprecipitates were assayed for autophosphorylation activity in 20 mM Tris-HCl, pH 7.5, 140 mM NaCl, 5 mM MnCl<sub>2</sub>, 10% glycerol,  $10 \mu\text{M}$  ATP and  $10 \mu\text{Ci}$  [ $\gamma$ -<sup>32</sup>P]ATP. The kinase-negative RafK375W (ref. 10) served as control. Samples were separated by SDS-PAGE and transferred to nitrocellulose. After exposure to X-ray film, blots were immunostained with anti-RafSP63 antiserum using the ECL kit (Amersham).



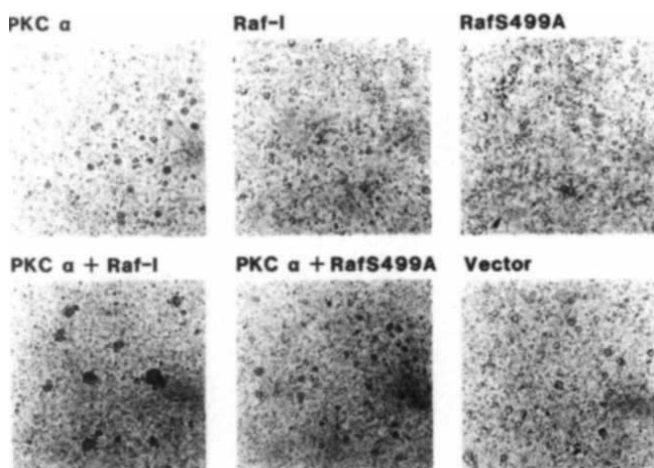


FIG. 4 PKC $\alpha$  synergizes with Raf-1 but not RafS499A to transform NIH3T3 cells. NIH3T3 were transfected with expression vectors as indicated and seeded in soft agar. Photographs were taken after 14 days.

17. Summers, M. D. & Smith, G. E. in *A Manual of Methods for Baculovirus and Insect Cell Culture Procedures* (Texas Agric. Exp. Station, College Station, 1987).
18. Kochs, G. et al. *Biochemical J.* (in the press).
19. App, H. et al. *Molec. cell. Biol.* **11**, 913–919 (1991).

ACKNOWLEDGEMENTS. W.K. and G.H. contributed equally to this work. We thank J. Knopf, H. Hug and B. Seed for PKC $\alpha$  cDNAs and expression vectors, and P. Lloyd, S. Hauser, C. Morath and I. Klamt for technical assistance.

## Induction temperature of human heat shock factor is reprogrammed in a *Drosophila* cell environment

Joachim Clos\*, Sridhar Rabindran\*, Jan Wisniewski & Carl Wu†

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Building 37, Rm 4C-09, Bethesda, Maryland 20892, USA

HEAT shock factor (HSF)<sup>1,2</sup>, the transcriptional activator of eukaryotic heat shock genes, is induced to bind DNA by a monomer to trimer transition involving leucine zipper interactions<sup>3,4</sup>. Although this mode of regulation is shared among many eukaryotic species, there is variation in the temperature at which HSF binding activity is induced. We investigated the basis of this variation by analysing the response of a human HSF expressed in *Drosophila* cells and *Drosophila* HSF expressed in human cells. We report here that the temperature that induces DNA binding and trimerization of human HSF in *Drosophila* was decreased by ~10 °C to the induction temperature for the host cell, whereas *Drosophila* HSF expressed in human cells was constitutively active. The results indicate that the activity of HSF *in vivo* is not a simple function of the absolute environmental temperature.

The DNA-binding activity of human HSF1 (hHSF1), the predominant of two related HSF proteins present in HeLa cells<sup>5,6</sup>, was analysed after transient expression in *Drosophila* SL2 cells. Little or no DNA-binding activity was detected in extracts from cells incubated at 22 °C or 27 °C (Fig. 1a). A substantial increase in activity was found at 32 °C, with maximal activity at 37 °C, normally the quiescent temperature for hHSF1. At 42 °C, when hHSF1 should be activated, its activity in *Drosophila* cells decreased abruptly. The activity of hHSF1 was resolved from endogenous *Drosophila* HSF (dHSF) by its faster electrophoretic mobility on nondenaturing gels, or by the removal of dHSF to the electrophoretic origin with specific polyclonal antibodies. Overall, the temperature response of hHSF1 was found to approximate closely the response of the endogenous dHSF in SL2 cells transfected with vector alone.

The changes in hHSF1 activity from 22 °C to 37 °C are not due to significant differences in the absolute amount of extracted protein. As shown by western blotting, equivalent levels of hHSF1 protein were observed at all temperatures, except for a sharp decrease at 42 °C (Fig. 1b), which may be caused by general insolubility or proteolysis. The induction of HSF DNA-binding activity involves a monomer to trimer transition that can be detected by gel filtration<sup>3,7,8</sup>. Similar to the gel filtration profile of hHSF1 expressed in human cells<sup>4</sup>, hHSF1 prepared from transfected *Drosophila* cells showed a characteristic increase in apparent molecular size (~200K to ~1,000K) upon heat shock at 37 °C (Fig. 1c).

\* Present addresses: Bernhard Nocht Institute for Tropical Medicine, Bernhard Nocht Strasse 74, D-2000 Hamburg, Germany (J.C.); American Cyanamid Co., Medical Research Division, 401 N. Middleton Rd., Pearl River, New York 10965, USA (S.R.).

† To whom correspondence should be addressed.

The activation of Raf-1 by coexpressed PKC $\alpha$  in Sf9 cells indicates that coexpression in NIH3T3 might result in transformation. Neither Raf-1 nor PKC $\alpha$  alone induced anchorage-independent growth. As anticipated, the combination of both enzymes supports the growth of small and medium-sized colonies in soft agar. RafS499A failed to synergize with PKC $\alpha$  in this assay (Fig. 4).

We conclude that phosphorylation of Ser499 is essential for efficient induction of Raf-1 kinase and transforming activity by PKC $\alpha$ . We further identified Ser259 as a second site required for PKC activation of Raf-1 autophosphorylation. Ser259 represents a Raf-1 autophosphorylation site<sup>14</sup>. Phosphorylation of Ser499 by PKC $\alpha$  is possibly a prerequisite for autophosphorylation of Ser259 and subsequent activation of Raf-1 by PKC $\alpha$ . In contrast, Ser43, another Raf-1 phosphorylation site (D. K. Morrison, G.H. and U.R.R., unpublished results), was not essential. These findings confirm and extend an earlier report of the activation of Raf-1 by PKCs  $\alpha$ ,  $\beta$  and  $\gamma$  in Sf9 cells and by PKC $\alpha$  *in vitro*<sup>15</sup>. Our data highlight the complexity of Raf-1 kinase regulation by covalent modification. Identification of PKC $\alpha$  as a Raf-1 kinase completes the cascade for mitogenic activation of oncogene class transcription factors, which has the sequence: PKC $\alpha$ →Raf-1→MAP kinase kinase→MAP kinase→TCF, Jun. The receptor tyrosine kinase activation of the cascade does not require PKC<sup>1,2</sup> and must involve a second mechanism of activation, which may include non-covalent modification. □

Received 9 November 1992; accepted 19 April 1993.

1. Rapp, U. R. *Oncogene* **6**, 495–500 (1991).
2. Heidecker, G., Kolch, W., Morrison, D. K. & Rapp, U. R. *Adv. Cancer Res.* **58**, 53–73 (1992).
3. Kyriakis, J. M. et al. *Nature* **358**, 417–421 (1992).
4. Dent, P. et al. *Science* **257**, 1404–1407 (1992).
5. Howe, L. R. et al. *Cell* **71**, 335–342 (1992).
6. Gille, H., Sharrrocks, A. D. & Shaw, P. E. *Nature* **358**, 414–417 (1992).
7. Pulverer, B. J., Kyriakis, J. M., Avruch, J., Nikolakakis, E. & Woodgett, J. *Nature* **353**, 670–674 (1991).
8. Smeal, T. et al. *Molec. cell. Biol.* **12**, 3507–3513 (1992).
9. Kolch, W. et al. *Oncogene* **5**, 713–720 (1990).
10. Heidecker, G. et al. *Molec. cell. Biol.* **10**, 2503–2512 (1990).
11. Piwnicka-Worms, H., Saunders, K. B., Roberts, T. M., Smith, A. E. & Cheng, S. H. *Cell* **49**, 75–82 (1987).
12. Thompson, P. A., Ledbetter, J. A., Rapp, U. R. & Bolen, J. B. *Cell Growth Different.* **2**, 609–617 (1991).
13. Williams, N. G., Roberts, T. M. & Li, P. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2922–2926 (1992).
14. McGrew, B. R. et al. *Oncogene* **7**, 33–42 (1992).
15. Sözeri, O. et al. *Oncogene* **7**, 2259–2262 (1992).
16. Luo, K., Hurley, T. R. & Sefton, B. M. *Meth. Enzym.* **201**, B 149–152 (1991).



We analysed the reversibility of hHSF1 binding in *Drosophila* cells in the presence of the protein synthesis inhibitor cycloheximide. HSF1 binding activity was observed to cycle in response to heat shock at 37 °C, recovery at 22 °C, and reinduction at 37 °C (Fig. 2a), in parallel with the cyclic changes in activity of the endogenous *Drosophila* factor<sup>9</sup>. The individual activity of

either HSF was confirmed by the inclusion of specific polyclonal antibodies in the gel shift assay. The antibody-dependent super-shift of HSF-DNA complexes also showed that hHSF1 and dHSF do not form mixed trimers, despite sequence similarity in the zipper motifs of the trimerization domain<sup>5,10</sup>. This finding excludes the possibility that hHSF1 trimerization was activated

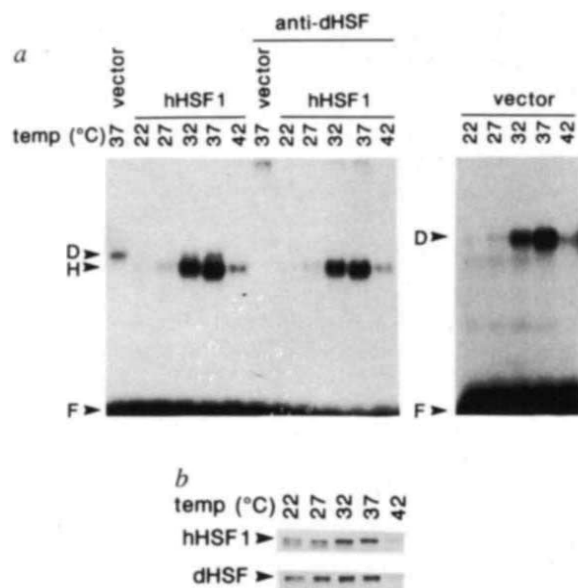


FIG. 1 Activity of hHSF1 in *Drosophila* SL2 cells. *a*, Gel mobility shift analysis of extracts from cells transfected with the expression vector or with the hHSF1 construct. Positions of free DNA (F) and DNA bound to hHSF1 (H) and the endogenous dHSF (D) are shown. Samples treated with antibodies to dHSF<sup>3</sup> (anti-dHSF) are indicated. The autoradiogram on the right was exposed 8 times longer than the one on the left. *b*, Western blot analysis of hHSF1 and endogenous dHSF in SL2 cells. The hHSF1 protein migrates as a cluster of closely spaced bands which coalesce at the heat shock temperatures, probably because of heat shock-induced modifications. *c*, Gel filtration chromatography of extracts prepared from control (22 °C) or heat-shocked (37 °C) cells transfected with the vector alone, and with constructs carrying dHSF and hHSF1. Arrows indicate molecular mass standards (thyroglobulin, 670K; gamma-globulin, 160K).

METHODS. DNA transfection with the plasmid for dHSF expression was as described<sup>4</sup>. To construct the hHSF1 expression plasmid, a PCR-amplified fragment containing nucleotides 1–1,590 (ref. 5) with 5'-AAT-TCAAA-3' added in front of the ATG codon was cloned into the *EcoRI*–*Bam*HI site of the pRmHa3 vector. After 16 h of induction with 0.7 mM CuSO<sub>4</sub>, cells were incubated for 15 min at the indicated temperatures and whole cell extracts were prepared as described<sup>9</sup> except that KCl was substituted for NaCl in extraction buffer. Gel mobility shift assays were done with 4 µg of total protein as described<sup>9</sup>. Where indicated, extracts were incubated with a 1:100 dilution of polyclonal rabbit antibody to dHSF<sup>4</sup> for 10 min before gel shift analysis. Total protein (8 µg) was analysed by western blotting. Nitrocellulose membranes were processed with rabbit polyclonal antibodies against hHSF1<sup>4</sup> (1:2,000) or dHSF<sup>3</sup> (1:1,000) as described<sup>4</sup>. Protein (30–40 µg; 50 µl) was used for gel filtration; fractions were analysed by SDS-PAGE (5% acrylamide for dHSF, 10% acrylamide for hHSF1), and by western blotting as described<sup>4</sup>.

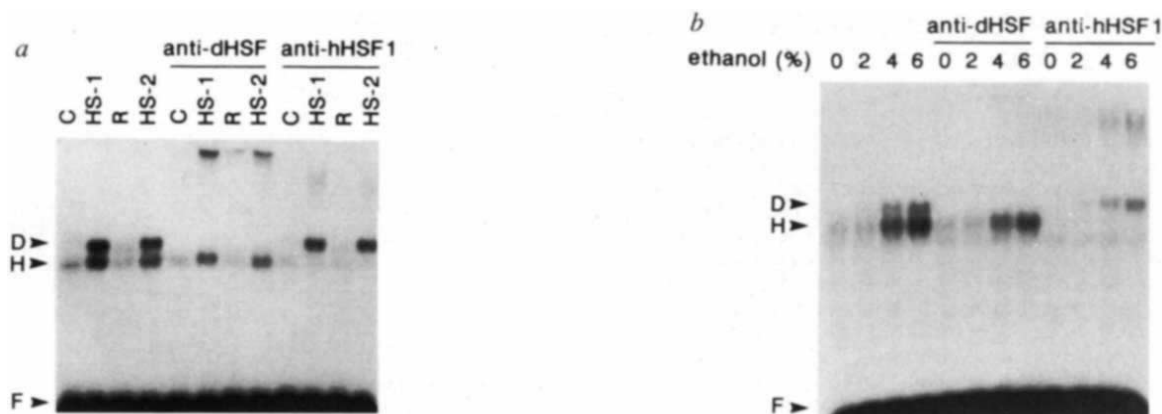


FIG. 2 *a*, Reversible activation of hHSF1 in *Drosophila* SL2 cells. Gel mobility shift analysis of extracts from SL2 cells expressing hHSF1 before heat shock (C), after the first heat shock (HS-1), recovery (R), and after a second heat shock (HS-2). After DNA transfection, cycloheximide (118 µM) was introduced 30 min before heat shock (15 min at 37 °C). Cells were allowed to recover at 22 °C for 2.5 h with vigorous aeration and heat shocked again. Identical samples of extract were also incubated

with antibodies specific for dHSF (anti-dHSF) and hHSF1 (anti-hHSF1). Free DNA (F) and DNA bound to the endogenous dHSF (D) and hHSF1 (H) are indicated. *b*, Induction of hHSF1 activity by ethanol treatment. Gel mobility shift analysis of extracts from SL2 cells expressing hHSF1. Cells were treated with ethanol at the indicated concentrations for 30 min. Extracts were pre-treated with antibodies to dHSF and hHSF1 as indicated.

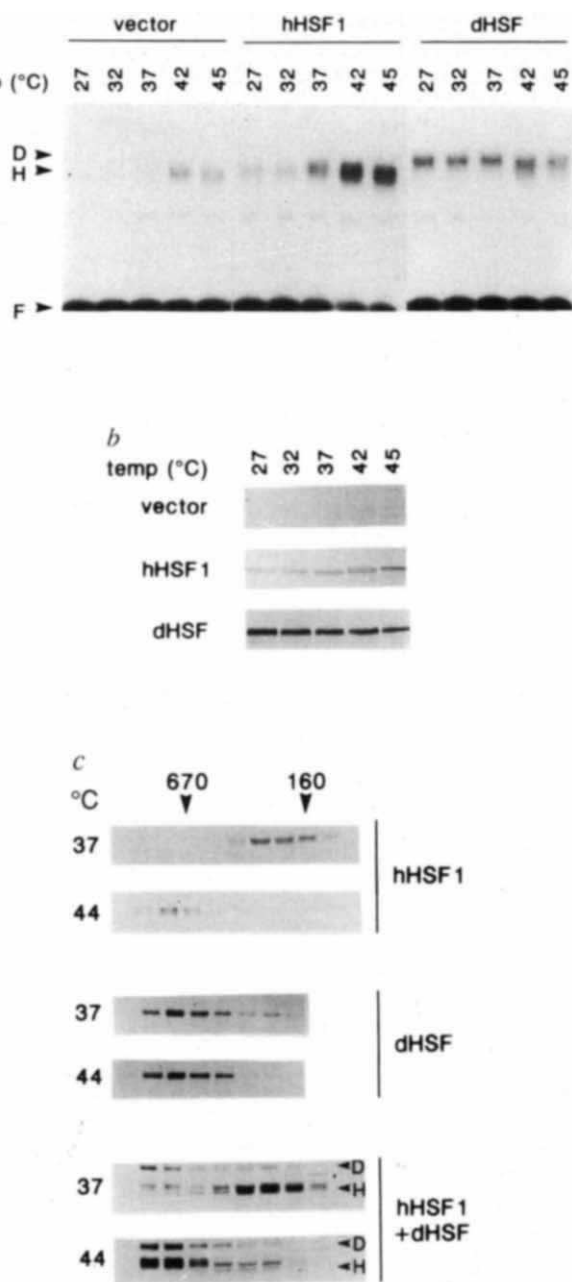


FIG. 3 Activity of hHSF1 and dHSF in human 293 cells. **a**, Gel mobility shift analysis of extracts prepared from cells transfected with the expression vector or with plasmids encoding hHSF1 or dHSF, and incubated for 30 min at the indicated temperatures. Free (F) DNA and DNA bound to hHSF1 (H) and dHSF (D) are indicated. **b**, Western blot analysis of expressed HSF proteins using antibodies specific for hHSF1 and dHSF. **c**, Gel filtration chromatography and western blotting of cell extracts prepared from control (37 °C) or heat-shocked (44 °C) cells. The upper panel shows endogenous hHSF1, detectable by increasing the extract concentration by five-fold; the lower panels show dHSF and hHSF1, as indicated. Molecular mass markers are indicated by arrows. **METHODS.** Plasmids for expression of hHSF1 and dHSF in human cells, and DNA transfection were as described<sup>4</sup>. The expression plasmid for dHSF was constructed by inserting the protein-coding region of pHSFWT<sup>10</sup> (XbaI-EcoRI fragment) in the XbaI-EcoRI site of pCMV5 (ref. 4). After transfection, cells were incubated at 37 °C for 14 h to allow expression of HSF proteins. For gel filtration, cells were collected and extracted immediately at the end of the 14 h expression period at 37 °C or after a heat shock for 20 min at the indicated temperatures. Gel mobility shift assays, western blots and gel filtration were as described in Fig. 1 legend.

at 37 °C in conjunction with dHSF merely by association with the zipper elements of dHSF. We also analysed the response of human HSF1 expressed in *Drosophila* cells to treatment with ethanol, a chemical inducer of the heat shock response, and found that the concentrations of ethanol (4–6%) required to induce dHSF also induced hHSF1 (Fig. 2b).

When dHSF was transiently expressed in human 293 cells, it was constitutively active at 37 °C and higher temperatures (Fig. 3a). The dHSF protein remained active even when transfected 293 cells were incubated at 32 °C, 27 °C (Fig. 3a) and 24 °C (data not shown), a procedure which slightly decreased the basal activity of the endogenous and transiently expressed human factor. Western blot analysis showed that equivalent amounts of *Drosophila* and human HSF proteins were present in extracts of 293 cells treated at the indicated temperatures (Fig. 3b). The dHSF protein chromatographed as a trimer (~1,000K)<sup>7</sup> by gel filtration when prepared from unshocked or heat shocked 293 cells, whereas endogenous hHSF1 formed trimers only upon temperature shift (Fig. 3c). Coexpression of both hHSF1 and dHSF in 293 cells did not alter the gel filtration profile of the expressed hHSF1 protein, excluding the possibility that the expression of dHSF alone was sufficient to induce the heat shock response in human cells.

A quantitative comparison of the DNA-binding activity of hHSF1 transiently expressed in human and in *Drosophila* cells shows that the threshold temperature for induction of the human protein is significantly lowered in *Drosophila* cells: at 32 °C, the binding activity is half-maximal whereas it remains at basal (~10%) levels in human cells (Fig. 4). These results indicate that the stability of the inactive HSF monomer can be substantially modulated by the cellular environment. The dependence on cellular factors in modulating HSF activity is also suggested from the constitutive trimerization of *Drosophila* HSF when expressed in human 293 cells. Despite its unregulated behaviour in human and *Escherichia coli* cells, dHSF does seem to exhibit a controlled response to heat shock when expressed in *Xenopus* oocytes<sup>10</sup> or in cultured *Drosophila* cells<sup>4</sup>.

Although the inducing temperature of the endogenous hHSF1 in HeLa cells can be decreased by ~1 °C when cells are grown

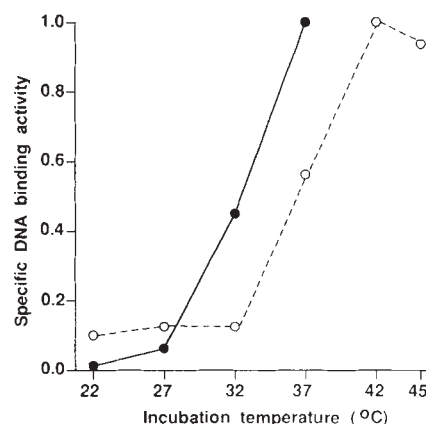


FIG. 4 Graph showing relative hHSF1 activity in extracts of transfected human 293 cells (open circles) and transfected *Drosophila* SL2 cells (filled circles) incubated at the indicated temperatures. Specific HSF binding activity was derived from values obtained by densitometer analysis of the autoradiograms shown in Figs 1a and 3a, and the western blots shown in Figs 1b and 3b. Values from human 293 cells were corrected for the endogenous hHSF1. The highest hHSF1 activity measured in each cell type was assigned a value of 1.0, and all other activities assigned relative values. The specific activity of hHSF1 in SL2 cells at 42 °C was not calculated, because of the inability to extract HSF (see text).



at 35 °C instead of 37 °C<sup>11</sup>, the greater reduction we observed for hHSF1 when it was expressed in *Drosophila* cells was surprising. This finding suggests that differences between human and *Drosophila* cell environments may lower the temperature threshold for induction by subtle alterations in the folding of the HSF1 monomer, or more likely, that destabilization of the latent hHSF1 protein is not strictly dictated by the absolute temperature, but by a change of temperature and the consequences of that change within the cell. Mechanisms for signalling intracellular changes in response to heat shock could be the temperature change itself or the ensuing decrease in the free pool of molecular chaperones, hsp70 in particular, which has been proposed to act

as a homeostatic regulator of HSF activity<sup>12,13</sup>. Although there is increasing evidence suggesting an interaction between HSF and hsp70 *in vitro*<sup>14,15</sup>, we have failed to observe a significant alteration in the induction of DNA-binding activity when intracellular concentrations of heat shock proteins are increased (S.R., J.W. and G. Li, unpublished observations). It is possible that other mechanisms, including post-translational modifications of HSF protein, also play an important role in regulating its activity in response to heat stress.

**Note added in proof:** The human HSF1 protein is also activated below its normal temperature threshold when expressed in tobacco protoplasts<sup>16</sup>. □

Received 19 January; accepted 29 April 1993.

1. Lis, J. T. & Wu, C. in *Transcription Book 2* (eds Yamamoto, K. R. & McKnight, S. L.) 907–930 (Cold Spring Harbor Laboratory Press, New York, 1992).
2. Sorger, P. *Cell* **65**, 363–366 (1991).
3. Westwood, J. T., Cios, J. & Wu, C. *Nature* **353**, 822–827 (1991).
4. Rabindran, S. K. *et al. Science* **259**, 230–234 (1993).
5. Rabindran, S. K., Giorgi, G., Cios, J. & Wu, C. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6906–6910 (1991).
6. Schuetz, T. J., Gallo, G. J., Sheldn, L., Tempst, P. & Kingston, R. E. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6911–6915 (1991).
7. Westwood, J. T. & Wu, C. *Molec. cell. Biol.* **13**, 3481–3486 (1993).
8. Perisic, O., Xiao, H. & Lis, J. T. *Cell* **59**, 797–806 (1989).
9. Zimarino, V. & Wu, C. *Nature* **327**, 727–730 (1987).
10. Cios, J. *et al. Cell* **63**, 1085–1097 (1991).

11. Abravaya, K., Phillips, B. & Morimoto, R. I. *Genes Dev.* **5**, 2117–2127 (1991).
12. Hightower, L. E. *Cell* **66**, 191–197 (1991).
13. Craig, E. A. & Gross, C. A. *Trends biochem. Sci.* **16**, 135–140 (1991).
14. Abravaya, K., Myers, M. P., Murphy, S. P. & Morimoto, R. I. *Genes Dev.* **6**, 1153–1164 (1992).
15. Baier, R., Welch, W. J. & Voellmy, R. *Cell Biol.* **117**, 1151–1159 (1992).
16. Treuter, E. *et al. Molec. gen. Genet.* (in the press).

**ACKNOWLEDGMENTS.** We thank B. Howard, W. Herr, T. Bunch and L. Goldstein for gifts of human and *Drosophila* expression vectors, T. Howard for advice on mammalian cell transfection, L. Sondergaard for the SL2 cell transfection protocol using Lipofectin, J. Eldridge for oligonucleotide synthesis and C. Klee for helpful comments. J.C., S.K.R. and J.W. made equal contributions to this work.

## Characterization of a functionally important mobile domain of GroES

Samuel J. Landry<sup>\*†</sup>, Jill Zeilstra-Ryalls<sup>†‡</sup>, Olivier Fayet<sup>†‡</sup>, Costa Georgopoulos<sup>‡§</sup> & Lila M. Gierasch<sup>\*</sup>

<sup>\*</sup>University of Texas Southwestern Medical Center, Dallas, Texas 75235-9041, USA

<sup>†</sup>Département de Biochimie Médicale, Université de Genève, 1211 Genève 4, Switzerland

<sup>‡</sup>University of Utah Medical Center, Salt Lake City, Utah 84132, USA

ALTHOUGH genetic<sup>1</sup> and biochemical<sup>2,3</sup> evidence has established that GroES is required for the full function of the molecular chaperone, GroEL, little is known about the molecular details of their interaction. GroES enhances the cooperativity of ATP binding and hydrolysis by GroEL (refs 4, 5) and is necessary for release and folding of several GroEL substrates<sup>6</sup>. Here we report that native GroES has a highly mobile and accessible polypeptide loop whose mobility and accessibility are lost upon formation of the GroES/GroEL complex. In addition, lesions present in eight independently isolated mutant *groES* alleles map in the mobile loop. Studies with synthetic peptides suggest that the loop binds in a hairpin conformation at a site on GroEL that is distinct from the substrate-binding site. Flexibility may be required in the mobile loops on the GroES seven-mer to allow them to bind simultaneously to sites on seven GroEL subunits, which may themselves be able to adopt different arrangements, and thus to modulate allosterically GroEL/substrate affinity.

The one-dimensional <sup>1</sup>H-nuclear magnetic resonance (<sup>1</sup>H-NMR) spectrum of GroES (Fig. 1a) contains sharp peaks superimposed on an envelope of broad, overlapping resonances. If GroES behaved as a roughly spherical folded protein oligomer of *M<sub>r</sub>* 70K, it would be expected to give broad NMR signals, consistent with an overall correlation time ( $\tau_c$ ) of ~50 ns; the sharp signals indicate that GroES has regions of polypeptide that are substantially more mobile ( $\tau_c$  < 1 ns). The

sharp peaks present in the one-dimensional spectrum lead to well-resolved crosspeaks in two-dimensional total correlated (TOCSY) and nuclear Overhauser effect (NOESY) spectra of GroES (Fig. 1b). Using methods of sequential assignment<sup>7</sup>, we assigned the resonances arising from the mobile segment of GroES to VETKSAGGIVLTGSAA, which spans amino acids 17–32 in the GroES sequence (Fig. 1c). As the assigned residues account for essentially all of the crosspeaks, flanking regions of GroES must be relatively immobile; hence we refer to this region as a mobile loop. The mobile loop region of GroES appears to lack any stable secondary structure on the basis of (1) the lack of NOEs diagnostic for specific structures<sup>7</sup> and (2) the striking similarity of all H<sup>α</sup> chemical shifts of loop residues with those of a synthetic peptide containing the loop sequence (see later) and with those reported for random-coil peptides<sup>8</sup>.

Limited proteolysis of GroES with trypsin produces fragments whose amino-terminal sequences indicate that there is a favoured cleavage site within the mobile loop at the K20–S21 peptide bond (Fig. 1d). We conclude that the mobile loop on GroES is accessible.

Among the 17 *groE* mutant strains originally isolated by their inability to support the growth of bacteriophage λ (ref. 9), eight were assigned to the *groES* gene using the ability of the λES<sup>+</sup>EL<sup>+</sup> transducing bacteriophage and its derivatives to propagate. All eight mutations in the *groES* structural gene occurred within the region spanning residues 23–31, namely just carboxy-terminal to the preferred trypsin cleavage site and coincident with the NMR-identified mobile region (Fig. 1c, and Table 1). On the basis of these phenotypes, we cannot distinguish between defective GroES function and increased turnover resulting in fewer copies of GroES protein. However, the following studies suggest that the mutant GroES proteins with altered mobile loop sequences are defective in their interaction with GroEL.

ATP-dependent formation of the GroES/EL complex causes substantial broadening of mobile loop NMR signals (Fig. 2c) and blocks the tryptic cleavage of GroES into ES1 and ES3 fragments (Fig. 2e). Thus the mobile loop becomes immobilized in the complex. The immobilization of the GroES loop could arise directly from an interaction with GroEL or indirectly from a GroEL-induced conformational change in GroES.

The possibility of a direct interaction of the GroES mobile loop with GroEL is supported by results with a synthetic

<sup>†</sup>Present addresses: Department of Biochemistry SL43, Tulane University School of Medicine, 1430 Tulane Avenue, New Orleans, Louisiana 70112 USA (S. J. L.); Department of Microbiology, University of Texas Medical School, 6431 Fannin, Houston, Texas 77225, USA (J. Z.-R.); Microbiology and Molecular Genetics-CNRS, 118 route de Narbonne, 31062 Toulouse, France (O. F.).

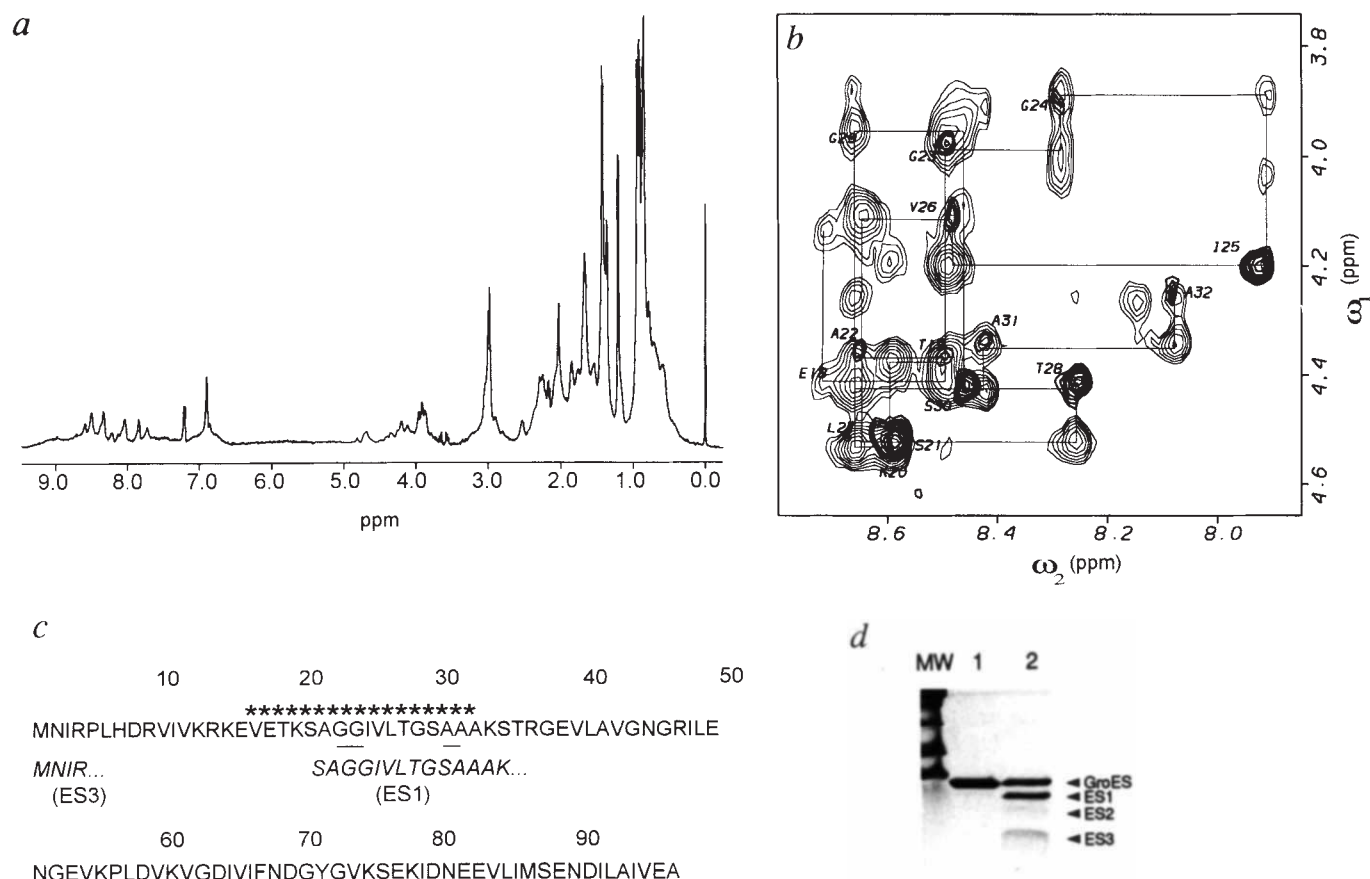


FIG. 1 Many resonance lines in the  $^1\text{H}$ -NMR spectrum of GroES are much narrower than expected for a 70K protein, indicating that a region of the protein is highly mobile (a). Sequential assignment of the mobile region using two-dimensional NMR spectra, TOCSY and NOESY (b, TOCSY in bold contours), reveals that the mobile region corresponds to residues 17–32 of GroES which include the trypsin-sensitive site (see below).  $^1\text{H}$  chemical shift assignments are as follows: V17, 4.17; E18, 4.43; T19, 4.39; K20, 4.50; S21, 4.53; A22, 4.33; G23, 3.99; G24, 3.91/4.03; I25, 4.22; V26, 4.14; L27, 4.51; T28, 4.42; G29, 4.09/4.00; S30, 4.45; A31, 4.36; A32, 4.28. The NMR-identified region of GroES is indicated in the GroES sequence<sup>21</sup> and is overlaid with asterisks (c). Residues changed in GroES mutants are underlined (see text and Table 1). GroES displays a proteolytically sensitive site in a highly mobile polypeptide loop. Limited proteolysis of GroES with trypsin produces two principal fragments, ES1 and ES3, whose size estimated from SDS-PAGE (d) and amino-terminal sequences (c) are as expected for cleavage at the K20-S21 peptide bond.

**METHODS.** Our earlier procedure<sup>10</sup> for purification of GroEL and GroES has been improved. Cultures of *E. coli* bearing plasmid pOF39 were grown to late log phase at 37 °C, collected by centrifugation at 4 °C, resuspended in 1/100th volume of ice-cold Tris-saline (50 mM Tris-HCl, 125 mM NaCl, pH 7.5) containing protease inhibitors (10  $\mu\text{M}$  E-64, 10  $\mu\text{M}$  aprotinin, 1 mM PMSF, 10 mM EGTA) and lysed by sonication. Subsequent purification steps up to but not including FPLC were carried out at 4 °C. The lysate was clarified by centrifugation at 25,000g for 20 min. and then fractionated by ammonium sulphate precipitation using solid ammonium sulphate. The 1.8–2.4 M ammonium sulphate cut was resuspended in Tris-saline and spun at 200,000g for 120 min. The supernatant was made 2.5 mM in  $\text{MgCl}_2$  and 1 mM in ATP and applied to a 5 $\times$ 30 cm Sephacryl S400HR gel-filtration column equilibrated with Tris-saline including  $\text{Mg-ATP}$ . By inspection of SDS-gels, approximately half of the GroES migrates as a complex with GroEL. Fractions containing GroEL/GroES complex were pooled, diluted 3-fold with distilled water and applied to a 30-ml Q-Sepharose column equilibrated with 20 mM bis-Tris-HCl, 50 mM KCl, pH 6.0. GroES and GroEL were eluted with a KCl gradient at  $\sim$ 0.2 and 0.5 M KCl, respectively. GroES and GroEL were each further purified by Mono-Q FPLC (Pharmacia) at room temperature using buffer and salt conditions as for the previous chromatographic step, except that GroEL was first passed over a 2-ml bed of cibacron-blue

agarose 3GA (Sigma). For each preparation of GroEL and GroES, the molar extinction coefficient at 280 nm was estimated by measuring the absorption of protein diluted in 50 mM Tris-HCl, 125 mM NaCl, 6 M guanidine-HCl, pH 7.5, where the stock protein solution was standardized by quantitative amino-acid analysis (Beckman System 6300). Typical  $A_{280}$  values with respect to GroEL and GroES protein subunits were  $14 \times 10^3$  and  $2.5 \times 10^3 \text{ mol}^{-1} \text{ cm}^{-1}$ , respectively. Both proteins were stored as the ammonium sulphate precipitate. Before an experiment, the protein was pelleted and washed in a Centricon-30 concentrator (Amicon) by 3 rounds of concentration and dilution with distilled water. All GroEL and GroES concentrations are expressed as the subunit concentration, determined from the absorbance at 280 nm. Limited proteolysis was carried out as follows: 50  $\mu\text{M}$  GroES was digested in 50  $\mu\text{l}$  10 mM Tris-HCl, pH 7.5 with 0.2  $\mu\text{g ml}^{-1}$  (d, lane 1) or 5  $\mu\text{g ml}^{-1}$  trypsin (d, lane 2) on ice for 30 min. Products were resolved by a 16.5% polyacrylamide gel using the Tricine-SDS buffer system<sup>22</sup> and stained with Coomassie blue. Molecular weight (MW) markers have the following  $M_s$ : 92,500, 66,200, 45,000, 31,000, 21,500 and 14,400 (Bio-Rad). An identical lane of unstained reaction products was electrophoretically transferred to polyvinylidene membrane and the bands visualized by staining with amido-black, cut out and sequenced (ABI model 477A) using the BLOTT-1 protocol (ABI User Bulletin 42). NMR methods were as follows: GroES samples were made  $\sim$ 1 mM (subunits) in an unbuffered solution near pH 7.0 containing 10%  $\text{D}_2\text{O}$  and 0.36 mM 3-(trimethylsilyl)propionate (TMSP). One-dimensional spectra were recorded at 25 °C on a Varian VXR500 NMR spectrometer and contained 16,000 complex points. Water suppression was accomplished by 1-s presaturation. 1 Hz line broadening was used in data processing. TOCSY and NOESY spectra were recorded at 10 °C with standard pulse schemes employing 'jump-and-return' pulses for water suppression. Mixing times were 40 ms for TOCSY and 150 ms for NOESY. For both TOCSY and NOESY, a total of  $2 \times 256$  FIDs of 1,024 complex points spanning 6,000 Hz were acquired with 32 scans per FID. Remaining water signal was removed from the FID numerically by subtracting a function generated from the moving average of a 16-point window<sup>23</sup>. Data were processed using the FTNMR software (Hare Research). FIDs in the  $t_1$  dimension were zero-filled to 1,024 points, and both dimensions were multiplied by a gaussian function. The chemical shift reference was trimethylsilyl of TMSP (0 p.p.m.).



TABLE 1 Phenotypes of *groES* mutant strains

| Mutant type | Mutation <sup>†</sup> |                          | Growth on L-agar* |     |     | Bacteriophage plaque formation <sup>†</sup> |     |
|-------------|-----------------------|--------------------------|-------------------|-----|-----|---------------------------------------------|-----|
|             |                       |                          | 37°               | 42° | 43° | λb2cl                                       | T5  |
| S-I         | G24>D                 | single copy <sup>§</sup> | +++               | ++  | —   | —                                           | —   |
|             |                       | multicopy <sup>‡</sup>   | +++               | +++ | +++ | +                                           | +/- |
| S-II        | G23>D                 | single copy              | +++               | +++ | ++  | —                                           | —   |
|             |                       | multicopy                | +++               | +++ | +/- | —                                           | +   |
| S-III       | Shine-Dalgarno, A31>V | single copy              | +++               | +++ | +++ | —                                           | +   |
|             |                       | multicopy                | +++               | +++ | +++ | +                                           | +   |

The *groES* alleles were cloned as follows: the *groE* mutant strains were marked with a kanamycin-resistance cassette immediately downstream of the *groE* operon. The *groES619* and *groES30* mutant alleles were then moved onto plasmids that contain chromosomal sequences flanking the *groE* operon but not the *groE* operon itself, by means of a RecA-mediated homologous recombination event. Such recombinant plasmids could be identified among plasmid DNA pools obtained from mutant strains bearing the parental *groE*-deletion plasmid by transforming wild-type *E. coli* cells with the DNA pools and selecting for resistance to kanamycin. The *groES42* allele was cloned from PCR-amplified DNA into pAPT110, a plasmid with a p15a origin of replication, in which expression of the *groES42* gene is under the control of the *lacUV5* promoter (details are provided in supplemental information). The other *groES* alleles were not cloned, as sequencing of PCR-amplified DNA showed them to be identical to the *groES619* allele.

\*Media also contained appropriate antibiotics. The symbol (+++) indicates growth identical to the wild-type *groE*<sup>+</sup> strain, in terms of both numbers of colony-forming units (CFU) and size of colonies. The symbol (++) indicates either 10-fold fewer CFUs or reduced colony size relative to the wild-type strain, or both. The symbol (+) indicates 10<sup>2</sup>-fold fewer CFUs (but no reduction in colony size) than the wild-type strain. The symbol (+/-) indicates 10<sup>3</sup>-fold fewer CFUs, and very small colony size relative to the wild-type strain. The symbol (—) indicates more than 10<sup>5</sup>-fold fewer CFUs than the wild-type strain.

<sup>†</sup>The symbol (+) indicates an efficiency of plating (EOP) of 1.0 and large plaque size. The symbol (+/-) indicates an EOP of 0.01–1.0 and small plaque size. The symbol (—) indicates lack of plaques (EOP < 10<sup>-6</sup>).

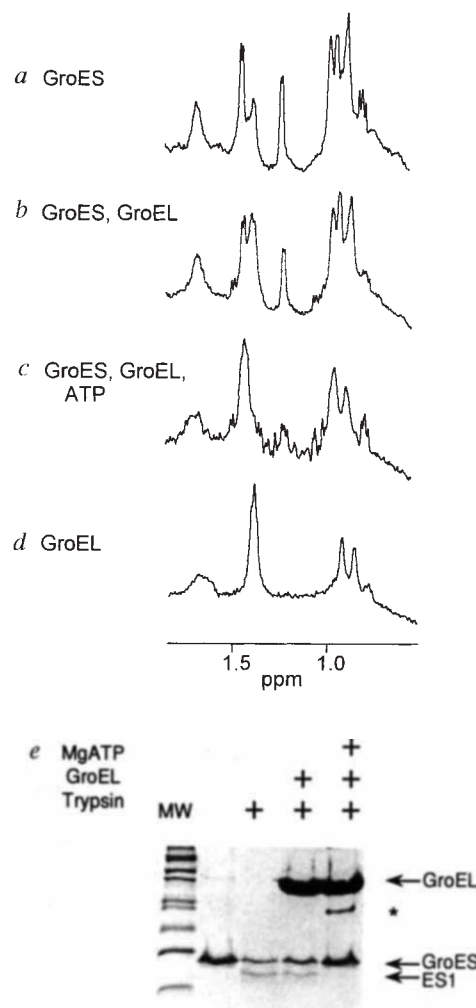
<sup>‡</sup>DNA sequences and allele names for eight *groES* mutants are as follows: GGC>GAC, S606, S3, SAC3, S619, S7, S97 (S-I); GGC>GAC, S42 (S-II); AAAGGAGAG>AAAGGAAAG (S-D) and GCA>GTA (A>V), S30 (S-III).

<sup>§</sup>The *groE* single-copy mutant strains bear the vector plasmid into which the *groE* allele has been cloned; for S-I and S-III this is pBR322, for S-II this is pAPT110 (see supplemental information).

||The *groE* multicopy mutant strains S-I and S-III were constructed by transforming the *groES* mutant bacteria with plasmid DNA bearing the cloned mutant *groE* alleles. An equivalent multicopy S-II strain was constructed by co-transforming the S-II mutant bacteria with the pAPT-110 plasmid derivative bearing the *groES42* mutant allele, plus a second plasmid bearing the wild-type *groEL* gene.

FIG. 2 The GroES mobile loop becomes immobilized and resistant to proteolysis in the complex with GroEL. <sup>1</sup>H resonances of the GroES mobile loop are essentially unaffected by GroEL alone but are severely broadened by GroEL and Mg-ATP (a–d). Note particularly the signal at 1.22 p.p.m. corresponding to the γ-protons of T19 and T28 of GroES which are substantially broadened in the complex (c). Appearance of GroES tryptic fragment ES1 indicates that the trypsin-sensitive site remains accessible in the presence of GroEL but not when Mg-ATP is included to form the GroES/GroEL complex (e). We note that a small amount of a tryptic fragment of GroEL appears in conditions where GroES is protected from tryptic digestion (asterisk in e, and data not shown). At present we do not know whether the GroEL fragment is more efficiently generated in the complex or preferentially protected by the complex.

METHODS. Limited proteolysis: 10 μM GroES was treated in 100 μl 10 mM Tris-HCl, 10 mM KCl, pH 7.5 with 5 μg ml<sup>-1</sup> trypsin in the presence of 20 μM GroEL, 2.5 mM MgCl<sub>2</sub> and 1 mM ATP (neutralized with KOH) as indicated. Reactions were analysed with a 10% polyacrylamide gel using the Tricine-SDS buffer system and stained with Coomassie blue. Molecular weight (MW) markers have the M<sub>r</sub>s: 200,000, 116,300, 97,400, 66,300, 55,400, 36,500, 31,000, 21,500, 14,400, 6,000 (Novex). NMR: Samples contained ~0.2 mM GroES and/or 0.4 mM GroEL and 2.5 mM MgCl<sub>2</sub>, 1 mM ATP, as indicated, in 80 mM d<sub>11</sub>-Tris-HCl, pH 7 (MSD Isotopes), 10% D<sub>2</sub>O, 0.36 mM TMS. Suppression of the water signal was by presaturation.



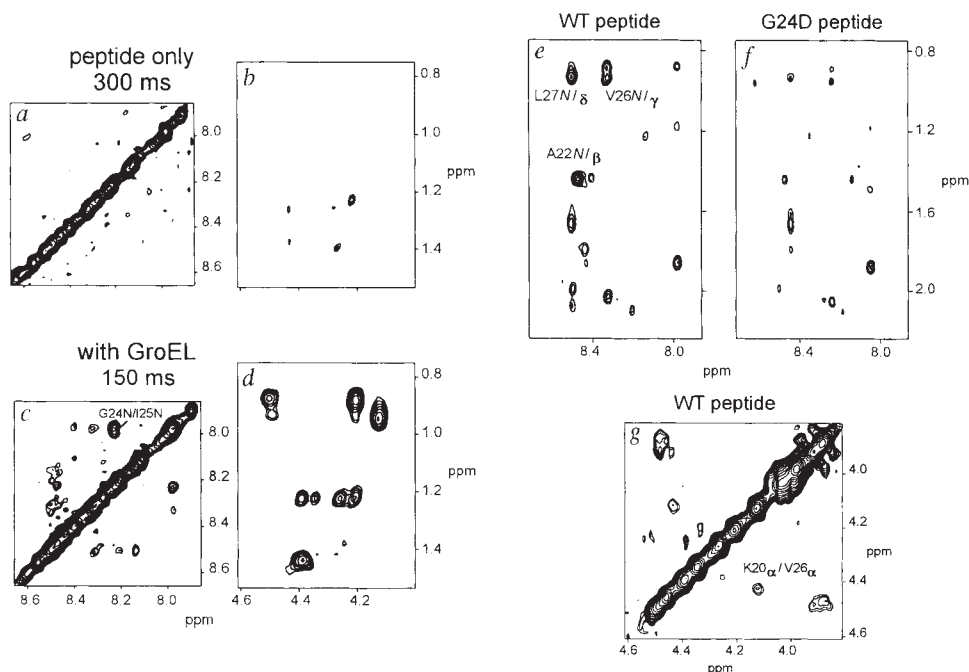
20-residue peptide containing the sequence of the loop KRK-EVETKSAGGIVLTGSAA. This peptide binds to GroEL as shown by the development of strong transferred NOEs (trNOEs) (Fig. 3c–e, g). In contrast to the wild-type peptide, a peptide corresponding to the G24>D mutant mobile loop exhibits smaller trNOEs (Fig. 3f). Some of the strongest trNOEs in the wild-type peptide occur in the sequence IVL, suggesting involvement of this region in binding. By contrast, the trNOEs observed for the mutant peptide are not localized to the hydrophobic cluster to the same extent, suggesting that it does not interact with GroEL in the same mode as the wild-type peptide.

We found previously (refs 10 and 11, and S.J.L., Z. Wang, J. Maxwell, T. Scott, and L.M.G., unpublished observations) that

FIG. 3 The wild-type mobile loop synthetic peptide (KRKEVETKSAG-GIVLTGSAA) binds to GroEL; whereas a peptide corresponding to the G24>D mutant GroES mobile loop (KRKEVETKSADGIVLTGSAA) binds less well. The NH/NH (a) and H<sup>α</sup>/aliphatic (b) regions of the NOESY spectrum of the wild-type peptide alone at room temperature show few NOEs because cross-relaxation is inefficient for the unstructured and rapidly tumbling peptide. The development of trNOEs in the peptide with GroEL indicates efficient cross-relaxation while bound to the slowly tumbling GroEL (c,d). The NH/aliphatic regions of NOESY spectra of wild-type (e) and G24>D mutant (f) peptides show that the mutant peptide binds to GroEL less tightly. Strong trNOEs in the wild-type spectrum, corresponding to protons in hydrophobic residues, are less prominent in the mutant, suggesting that the mode of binding by the mutant is less localized to the hydrophobic cluster. The H<sup>α</sup>/H<sup>α</sup> region

of the spectrum of wild-type peptide with GroEL shows a long-range trNOE, which indicates that a turn is present in the GroEL-bound conformation (g).

**METHODS.** Peptides were synthesized with a Milligen 9050 continuous flow peptide synthesizer using 9-fluorenyl-methoxycarbonyl (FMOC) chemistry and purified by reversed-phase HPLC. NMR samples contained 4 mM peptide (determined by quantitative amino-acid analysis) and 120 μM GroEL (c-e) in 40 mM potassium phosphate buffer, pH 6.0 with 10% D<sub>2</sub>O and 0.36 mM TMSP. Resonance



assignments were obtained from TOCSY and NOESY spectra taken at 8 °C using mixing times of 65 and 300 ms, respectively. H<sup>α</sup> chemical shift assignments for the wild-type peptide using GroES sequence numbering are as follows: V17, 4.15; E18, 4.42; T19, 4.36; K20, 4.44; S21, 4.50; A22, 4.32; G23, 4.00; G24, 3.99; I25, 4.22; V26, 4.14; L27, 4.52; T28, 4.41; G29, 4.05; S30, 4.48; A31, 4.36; A32, 4.28. For the trNOE experiments, NOESY spectra were recorded at 25 °C, with mixing time 150 ms.

many peptides bind to GroEL, and that their binding correlates with their ability to present a hydrophobic face. We proposed that this mode of binding mimics substrate binding to GroEL. Competition studies argue that the GroES mobile loop peptide binds to a distinct site, not coincident with that of substrate binding. Whereas peptides that bind through their hydrophobic face compete with each other for GroEL binding (as indicated by reduced trNOE size on the target peptide in the presence of excess competitor), the mobile loop peptide develops trNOEs of similar size in the presence or absence of excess substrate-mimic peptides (data not shown). By contrast, addition of GroES plus ATP substantially reduces the trNOEs of the mobile loop peptide, suggesting that some of the peptide binding sites are blocked by formation of the GroES/GroEL complex. Complete loss of loop peptide trNOEs may not be expected as GroES binds to only one GroEL toroid at a time<sup>12</sup>.

A number of trNOEs in the GroES mobile loop peptide, including K20H<sup>α</sup>/V26H<sup>α</sup>, S21H<sup>β</sup>/L27H<sup>β</sup> and G24NH/I25NH (Fig. 3, and data not shown), indicate that the mobile loop peptide binds to GroEL in a hairpin conformation with a turn occurring in the sequence, S21-A22-G23-G24-I25 (using GroES numbering). The hairpin conformation induced in the loop peptide upon interaction with GroEL stands in contrast to the conformations induced in substrate-mimic peptides, where clustering of hydrophobic groups appears to be the primary conformational determinant and often helices are induced upon binding<sup>10, 11</sup>. Furthermore, sequence comparison of GroES homologues<sup>13</sup> (many of which can interact productively with *E. coli* GroEL (refs 14–16)), reveals a high degree of conservation of the mobile loop region, in particular of these small polar residues, which may be imposed by binding requirements. Taken together, the competition results, the specific bound conformation, and the sequence comparisons argue that the mobile loop region of GroES binds to GroEL at a GroES-specific site.

Ultrastructural results<sup>12,24</sup> favour a model for GroEL-substrate interactions in which the substrate is sequestered within the GroEL double-doughnut<sup>17–20</sup>. Under the influence of ATP, changes in GroEL/substrate affinity lead to release of segments of the substrate, which in the absence of successful folding, either rebinding or aggregate. GroES can be viewed in this model as a mediator of GroEL subunit-subunit interactions. Concomitant binding of seven GroES subunits to seven GroEL subunits could coordinate the conformational changes in GroEL in order to release substrate from all GroEL sites simultaneously. Mobility of the GroES binding sites could enable them to accommodate substrate-specific distortions in the GroEL tetradecamer. □

Received 11 December 1992; accepted 27 April 1993.

1. Tilly, K. & Georgopoulos, C. *J. Bact.* **149**, 1082–1088 (1982).
2. Chandrasekhar, G. N. *et al. J. Biol. Chem.* **261**, 12414–12419 (1986).
3. Goloubinoff, P. *et al. Nature* **342**, 884–889 (1989).
4. Gray, T. E. & Fersht, A. R. *FEBS Lett.* **292**, 254–258 (1991).
5. Bochkareva, E. S. *et al. J. Biol. Chem.* **267**, 6796–6800 (1992).
6. Jaenicke, R. *Curr. Opin. Struct. Biol.* **3**, 104–112 (1993).
7. Wuthrich, K. *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).
8. Wishart, D. S., Sykes, B. D. & Richards, F. M. *J. molec. Biol.* **222**, 311–333 (1991).
9. Georgopoulos, C. P. *et al. J. molec. Biol.* **76**, 45–60 (1973).
10. Landry, S. J. & Gierasch, L. M. *Biochemistry* **30**, 7359–7362 (1991).
11. Landry, S. J. *et al. Nature* **355**, 455–457 (1992).
12. Langer, T. *et al. EMBO J.* **11**, 4757–4765 (1992).
13. Hartman, D. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 3394–3398 (1992).
14. Lubben, T. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 7683–7687 (1990).
15. Bertsch, U. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 8696–8700 (1992).
16. Ohtaka, C., Nakamura, H. & Ishikawa, H. *J. Bact.* **174**, 1869–1874 (1992).
17. Martin, J. *et al. Nature* **352**, 36–42 (1991).
18. Creighton, T. E. *Nature* **352**, 17–18 (1991).
19. Mendoza, J. A. *et al. J. Biol. Chem.* **266**, 13044–13049 (1991).
20. Nilsson, B. & Anderson, S. A. *Rev. Microbiol.* **45**, 607–635 (1991).
21. Hemmingsen, S. M. *et al. Nature* **333**, 330–334 (1988).
22. Schagger, H. & von Jagow, G. *Analyt. Biochem.* **166**, 368–379 (1987).
23. Marion, D., Ikura, M. & Bax, A. *J. magn. Reson.* **84**, 425–430 (1989).
24. Braig, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **90**, 3978–3982 (1993).

**ACKNOWLEDGEMENTS.** We thank N. Nguyen for preliminary work and M. Briggs for discussion. This work was supported by NIH and FNS.



# PAS is a dimerization domain common to *Drosophila* Period and several transcription factors

Zuoshi Josh Huang, Isaac Edery & Michael Rosbash\*

Howard Hughes Medical Institute, Department of Biology, Brandeis University, Waltham, Massachusetts 02254, USA

\* To whom correspondence should be addressed.

MUTATIONS in the *period* gene product (PER) can shorten or lengthen the circadian rhythms of *Drosophila melanogaster*<sup>1</sup>, but its biochemical activity has not been established. PER contains a motif of ~270 amino acids whose function is unknown (termed PAS<sup>2</sup>) and which is also present in three transcription factors of the basic-helix-loop-helix (bHLH) type, in the *D. melanogaster* single-minded gene product (SIM)<sup>2</sup>, and in both subunits of the mammalian dioxin receptor complex<sup>3,4,28</sup>. We show here that the PER PAS functions *in vitro* as a novel protein dimerization motif and that it can mediate associations between different members of the PAS protein family. The dimerization efficiency is decreased by several missense mutations in the PAS domain, including the original *per*<sup>L</sup> mutation, which lengthens circadian periods from 24 h to 29 h (ref. 1). The results indicate that the PAS domain may function as a dimerization domain in both SIM and the dioxin receptor complex, and that PER may regulate circadian gene transcription partly by interacting with the PAS domain of bHLH-PAS-containing transcription factors.

To test whether the PAS domain can mediate protein-protein interactions, we designed a co-immunoprecipitation assay: PER fragments containing complete or truncated PAS regions were

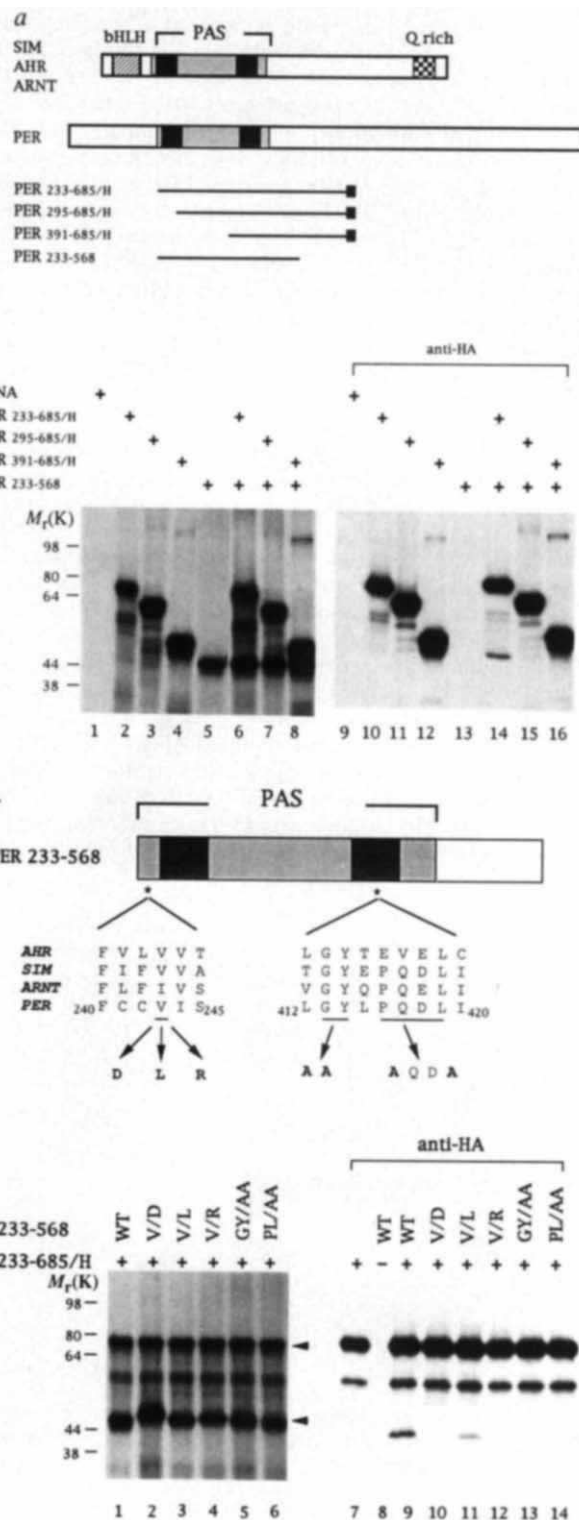


FIG. 1 The PAS domain mediates PER protein self-association *in vitro*. **a** (top), Schematic representation of the bHLH-PAS proteins (SIM, AHR and ARNT) and the full-length PER protein. bHLH motif, PAS domain, the two 51-amino-acid direct repeats (here termed PAS repeats) (dark stippling), and glutamine-rich (Q-rich) region are indicated; bottom, *in vitro* synthesized PAS-containing PER fragments with and without the haemagglutinin epitope (HA-tag, black square). Numbers represent the first and last amino acids of the PER fragment. **b**, Single or mixed *in vitro* translation products were analysed on a 12% SDS-polyacrylamide gel, either directly (lanes 1–8) or after immunoprecipitation with monoclonal antibody 12CA5 against the HA-tag (anti-HA, lanes 9–16). The co-immunoprecipitated fragment (PER 233–568) is indicated by an arrow. Smaller minor products probably result from internal translation initiation or from degradation. **c**, Wild-type and 5 mutants of PER 233–568 fragment. Two stretches of amino acids that are highly conserved among PAS family members are shown below the asterisks. The first and the last amino acids of each stretch in the PER protein are numbered. Mutated amino acids are underlined, changes being shown in bold letters under the arrows and named, from left to right, V/D, V/L, V/R, GY/AA and PL/AA. **d**, Single or mixed (with PER 233–685/H) *in vitro* translation products analysed by 12% SDS-PAGE either directly (lanes 1–6) or after immunoprecipitation with anti-HA antibody (lanes 7–14). The migration positions of PER 233–685/H (top arrowheads) and different versions of PER 233–568 (lower arrowheads) are indicated; the ~55K translation product probably results from internal initiation or degradation of PER 233–685/H. The residual co-precipitation of the V/D fragment (lane 10) is not evident here but was visible in the original autoradiograph.

**METHODS.** *per* cDNA encoding PER 233–685/H was generated by the polymerase chain reaction (PCR) from full-length *per* cDNA (pSP65ATper; ref. 22). The 5' primer contains, from 5' to 3', an XbaI site, an ATG codon and nucleotides 697 to 717 of *per* cDNA (nucleotide 1 is referred to as the start of translation); the 3' primer contains, from 5' to 3', an XbaI site, a TAG stop codon, the sequence corresponding to the HA peptide (YPYDVPDVASL; ref. 23), and nucleotides 2,052 to 2,035 of *per* cDNA. PCR products were inserted into the pBluescript KS (–) vector (Stratagene) at the XbaI site, such that the *per* cDNA was under the control of the T7 promoter (pBSC2H), and was verified by DNA sequencing. cDNAs coding for PER 295–685/H and PER 391–685/H were obtained by PCR using pBSC2H as a template. The 5' primers contain a T7 promoter, the 5' leader and ATG from the human  $\beta$ -globin gene<sup>24</sup>, and nucleotides 883 to 902 (PER 295–685/H), or nucleotides 1,174 to 1,193 (PER 391–685/H) of the *per* cDNA; the 3' primer for both cDNAs contains nucleotides 806 to 787 of the (+) strand of the pBluescript (KS–) vector. The cDNA for PER 233–568 was generated by linearizing pBSC2H with *Sma*I at nucleotide 1,704 of *per* cDNA. Site-directed mutagenesis (Amersham kit) was carried out with

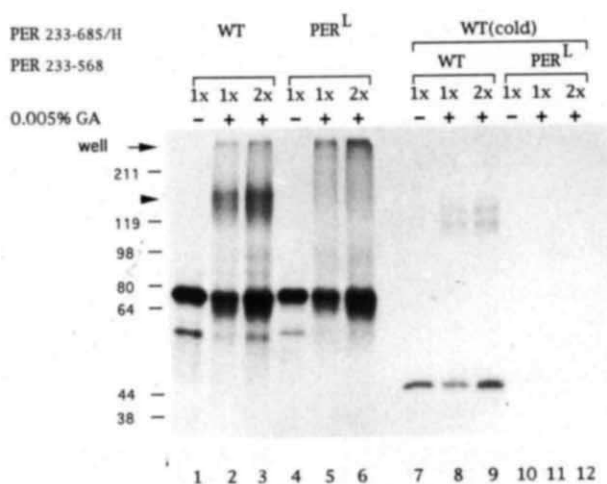
pBSC2H as template and verified by sequencing. These mutant pBSC2H templates were linearized with *Sma*I to obtain cDNAs for mutant PER 233–568 fragments. *In vitro* transcription and translation followed the manufacturer's instructions (Promega). <sup>35</sup>S-methionine-labelled rabbit reticulocyte lysate translation products giving comparable c.p.m. were mixed and incubated at 37°C for 30 min and 1  $\mu$ l analysed by SDS-PAGE; for immunoprecipitation, 5  $\mu$ l was typically diluted in 250  $\mu$ l ice-cold HND buffer (20 mM HEPES, 100 mM KCl, 10% glycerol, 0.4% NP-40, 5 mM EGTA, 5 mM EDTA, 100  $\mu$ g ml<sup>–1</sup> BSA, 1 mM DTT, pH 7.4). After pre-clearing with 7  $\mu$ l Gamma-Bind Plus Sepharose (Pharmacia) and 10 min of rocking, mixtures were spun and the supernatant incubated with 2  $\mu$ l monoclonal antibody 12CA5 (Babco) for 2 h. Gamma-Bind Plus Sepharose (10  $\mu$ l) was then added with further incubation for 1 h, samples were washed 3 times with HND buffer, heated at 95°C for 5 min in SDS sample buffer, centrifuged for 2 min, and analysed by 12% SDS-PAGE. Gels were fixed, amplified, fluorographed and quantified by phosphorimager scanning.

synthesized *in vitro*, either with or without a haemagglutinin epitope (HA tag) at the C terminus (Fig. 1a). Without an HA tag, PER 233–568 was not immunoprecipitated by the anti-HA antibody 12CA5 (Fig. 1b: compare lanes 10–12 with lane 13). It was, however, co-immunoprecipitated after mixing with PER 233–685/H (Fig. 1b, lane 14), indicating that the two polypeptides associate *in vitro*. Between 10 and 15% of PER 233–568 co-precipitated with PER 233–685/H when they were mixed in ~1:1 molar ratio (based on four experiments). This is an underestimate of PER self-association as PER 233–568/PER 233–685 associations are not detectable in the assay, and the association conditions may not be optimal. Full-length PER protein also associated with PER 233–685/H (data not shown). Significantly, the co-immunoprecipitation of PER 233–568 with PAS deletion fragments was greatly reduced (Fig. 1b, lanes 15 and 16).

Conserved amino acids in the PAS domain were then mutated and the association of the mutant polypeptides with wild-type PER fragments assayed (Fig. 1c and d). Of particular interest was the valine at position 243 (V243), which lies in a conserved hydrophobic region just N-terminal to the first PAS repeat (Fig. 1c). In the classical *per<sup>L</sup>* mutation (a valine to an aspartic acid (V243D), missense mutation at position 243; refs 1,5), the protein (PERL) is biologically active but gives rise to 29-h circadian rhythms. The sequence of PER 233–568 was modified to change V243 to aspartic acid as well as to leucine and arginine (Fig. 1c). Another highly conserved stretch of amino acids in the second PAS repeat was also mutated (Fig. 1c). With one exception, all of the mutations resulted in a significant reduction in co-immunoprecipitation with wild-type PER 233–685/H (Fig. 1d, lanes 9–14). The original *per<sup>L</sup>* (V/D) mutation showed the most severely affected phenotype (Fig. 1d, compare lanes 9 and

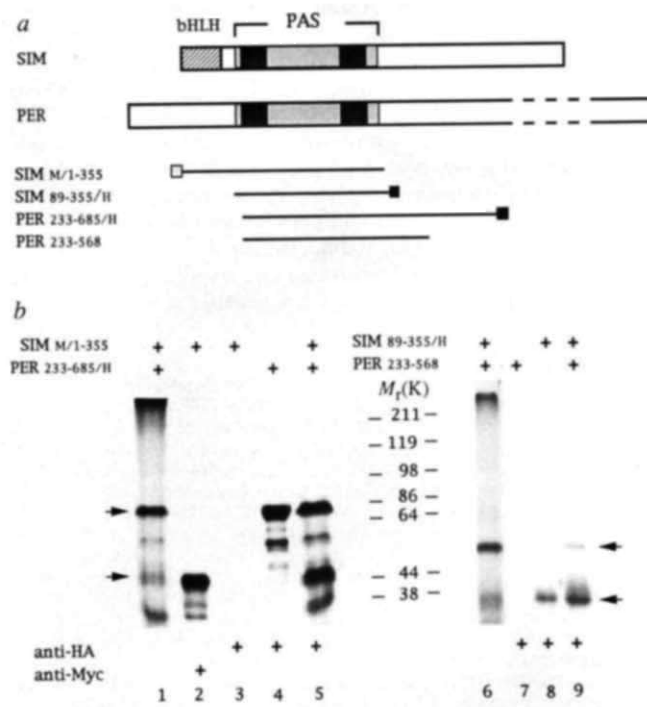
10), a ~7-fold reduction in association with PER 233–685/H (based on three experiments; see Fig. 1d legend). It also resulted in a slower migration on SDS-polyacrylamide gels owing to an effect of charge or a structural change in the protein fragment (Fig. 1d, compare lanes 2 and 1). A further decrease in PER self-association was observed when both PER fragments carried the *per<sup>L</sup>* mutation (data not shown; also see Fig. 2). The conservative change of valine to leucine at position 243 (V/L) resulted in only a mild reduction in co-immunoprecipitation (Fig. 1d, lane 11). These observations suggest that the PAS domain engages in PAS-PAS associations.

To study the stoichiometry of the PER protein oligomers *in vitro*, we designed a chemical crosslinking and immunoprecipitation procedure. After crosslinking with glutaraldehyde and immunoprecipitation with anti-HA antibody, the wild-type fragment but not the PERL mutant fragment gave rise to dimer-sized crosslinked products that migrated as a doublet at  $M_r \sim 140$ –160K (Fig. 2, compare lanes 2 and 3 to lanes 5 and 6: the doublet is indicated by an arrowhead on the left; monomers are ~75K polypeptides, see lanes 1 and 4). This doublet probably reflects crosslinking at two different locations, giving



**FIG. 2** PER forms dimers *in vitro*. Translations of  $^{35}\text{S}$ -labelled wild type (lanes 1–3) or *per<sup>L</sup>* version (lanes 4–6) of PER 233–685/H were incubated in the presence (+) or absence (–) of 0.005% glutaraldehyde (GA) in phosphate buffer. They were then immunoprecipitated with 12CA5 and analysed on a 10% SDS-polyacrylamide gel. Arrowhead (left) indicates crosslinked homodimers of wild-type (WT) PER 233–685/H. In lanes 7–12, unlabelled PER 233–685/H was mixed with the same number of c.p.m. of  $^{35}\text{S}$ -labelled wild-type (lanes 7–9) or *per<sup>L</sup>* version (lanes 10–12) of PER 233–568. They were then crosslinked and immunoprecipitated. Arrowhead (right) indicates crosslinked 'heterodimers' between cold (unlabelled) PER 233–685/H and radiolabelled PER 233–568. Amount of  $^{35}\text{S}$ -labelled products used was '1X' or '2X'. The origin of the gel is indicated (well).

**METHODS.** For lanes 1–6, 2  $\mu\text{l}$   $^{35}\text{S}$ -methionine-labelled PER 233–685/H was incubated in 40  $\mu\text{l}$  0.1 M phosphate buffer, pH 7.4, and GA at 25°C for 30 min; crosslinking was terminated by adding 2M Tris, pH 7.5, to 100 mM. Samples were diluted into 250  $\mu\text{l}$  HND buffer and immunoprecipitated as for Fig. 1. In lanes 7–12, unlabelled PER 233–685/H was mixed with  $^{35}\text{S}$ -labelled PER 233–568 at 37°C for 30 min; crosslinked and immunoprecipitated.



**FIG. 3** The PAS domain mediates associations between PER and SIM. **a**, SIM and PER aligned according to their PAS domains: PAS and bHLH domains are indicated. Below, SIM and PER fragments with a Myc tag (stippled square) or HA tag (black square) are shown. **b**, Single or mixed translation products of SIM and PER fragments were analysed by 10% SDS-PAGE, either directly (lanes 1 and 6) or after immunoprecipitation with anti-Myc or anti-HA antibody. Arrows (left) indicate PER 233–685/H (upper) and SIM M/1–355 (lower); on the right, PER 233–568 (upper) and SIM 89–355/H (lower). SIM M/1–355 is barely detectable in lane 1.

**METHODS.** Procedures as described for Fig. 1. cDNAs encoding SIM M/1–355 and 89–355/H were made by PCR from pNB40sim (gift from S. Crews). The 5' primer for SIM M/1–355 contains an *Xba*I site, an ATG codon, nucleotides encoding the Myc peptide (EQKLISEEDLN; ref. 25), and nucleotides 641–660 of *sim* cDNA; 3' primer has an *Xba*I site, TAG stop codon, and nucleotides 1,708–1,690 of *sim* cDNA. The PCR product was inserted at the *Xba*I site of the pBluescript KS (–) so that the *sim* cDNA was under the control of the T7 promoter. The 5' primer for SIM 89–355/H contains a T7 promoter, the 5' leader and ATG from the human  $\beta$ -globin gene, and nucleotides 888–909 of the *sim* cDNA; 3' primer contains a stop codon, HA peptide sequence, and nucleotides 1,708–1,690.



rise to different covalent dimers with different cross-sectional areas. To confirm that two PAS-containing proteins can dimerize, an additional crosslinking-immunoprecipitation experiment was done in which non-radioactive PER 233–685/H was mixed with the smaller, 45K radiolabelled PER 233–568 (either the wild-type or the PERL version). As the assay only detects radioactive oligomers that also contain an HA epitope, dimer formation must result from an association between a non-radioactive HA-containing subunit and a radiolabelled subunit. An appropriately sized doublet (about 110–130K; indicated by the arrowhead on the right in Fig. 2) was detected only with the wild-type PER fragment (Fig. 2; compare lanes 8 and 9 to 11 and 12), confirming that the PAS domain can dimerize *in vitro* and that the PERL fragment is at least ten times less efficient than wild-type in this assay. Dimer formation does not rule out the presence of small amounts of larger oligomers.

In contrast to its three relatives, PER has no known DNA-binding domain or DNA-binding activity (data not shown). As there is evidence that PER affects circadian transcriptional regulation<sup>6–8</sup>, it may participate in interactions with other PAS-containing DNA-binding proteins. We therefore tested whether PER can associate with other members of the PAS family that contain DNA-binding regions. A SIM fragment containing the bHLH and PAS domain with a Myc epitope at its N terminus (SIM M/1–355; Fig. 3a) was immunoprecipitated by the anti-Myc antibody 9E10, but not by the anti-HA antibody 12CA5 (Fig. 3b, lanes 2 and 3). When mixed with PER 233–685/H, it was co-precipitated by the anti-HA reagent (Fig. 3b, lanes 1 and 5), indicating that the two fragments can associate *in vitro*. SIM PAS was sufficient for this association because PER 233–568 was co-precipitated with a SIM fragment containing an HA-tag and only the PAS domain (SIM 89–355/H; Fig. 3b, lanes 6–9). The PER PAS domain can also mediate a similar association between protein fragments of PER and the aryl hydrocarbon nuclear translocator (ARNT), a subunit of the dioxin receptor complex (data not shown). Taken together, the observations suggest that PAS may contribute to heterodimer formation between different PAS-containing family members, for example between the two dioxin receptor

subunits, which are known to function as a heterodimer.

To investigate whether PER can undergo homotypic interactions *in vivo*, we crossed two transformant lines bearing two distinct *per*  $\beta$ -galactosidase (*per*- $\beta$ -gal) constructs (SG and NG in Fig. 4a) to flies carrying a biologically active, epitope-tagged version of PER (HA/PER)<sup>9</sup>. The entrained progeny were collected at Zeitgeber time 22 (ZT22), when PER levels are maximal<sup>10</sup>, total head extracts immunoprecipitated with anti-HA antibodies and  $\beta$ -galactosidase activity measured in the immune complexes (Fig. 4b).

Our results show that HA/PER forms stable complexes with the SG fusion protein (SG×HAper in Fig. 4b) and that PER may undergo homotypic interactions *in vivo*. Control experiments using either non-tagged *per* flies crossed to SG flies, or HAper flies crossed to NG flies (NG×HAper) show that immunoprecipitation of  $\beta$ -galactosidase activity depends on the presence of both the HA epitope and PER sequences in the N-terminal 630 amino acids of the SG fusion protein (Fig. 4a). The control experiments also argue that this pseudo-homotypic interaction occurs *in vivo* and not *in vitro*: immunoprecipitation of an extract made from mixed SG and HAper heads (SG+HAper) resulted in no  $\beta$ -galactosidase activity above negative control values. Although the results are consistent with the existence of an *in vivo* PAS–PAS interaction, the PAS region comprises only ~40% of the *per* coding region in SG (Fig. 4a). Therefore we cannot rule out the contribution of non-PAS sequences to the *in vivo* interaction.

In the three known bHLH–PAS proteins, the PAS domains are located just C-terminal to the bHLH motifs, like the leucine zipper (zip) domain in the bHLH–zip protein family (as in Myc and Max for example)<sup>11</sup>. The bHLH–PAS proteins may thus represent a new subfamily of bHLH proteins in which the PAS domain's role is similar to that of the leucine zipper in bHLH–zip proteins. As the dioxin-binding region of the aryl hydrocarbon receptor (AHR) is part of a PAS domain<sup>4</sup>, ligand binding may be connected with PAS-mediated protein–protein interaction (that is, heterodimer formation between AHR and ARNT, the two subunits of the dioxin receptor complex), not unlike the behaviour of some steroid hormone receptors<sup>12,13</sup>. Unknown

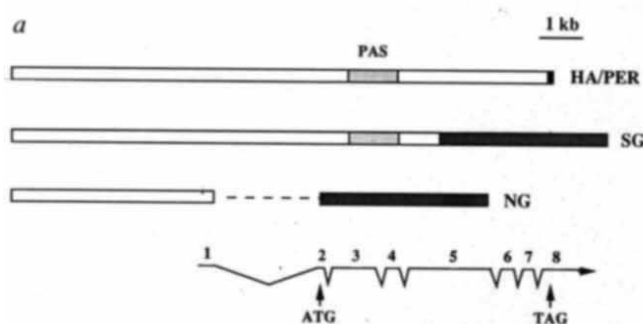
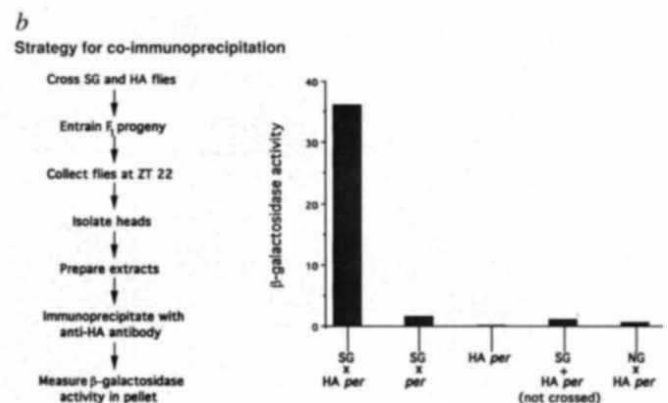


FIG. 4 PER can undergo homotypic interactions *in vivo*. *a*, Plasmid constructs used to generate HA/PER, SG and NG transformant flies<sup>9,26,27</sup>. Bars represent constructs containing *per* DNA (open),  $\beta$ -galactosidase DNA (hatched) or haemagglutinin epitope DNA (black), inserted into the cp20.1 P-element transformation vector. Dashed line indicates deleted first intron sequences; stippled box, PAS domain. Exons (1–8) and introns of the principal form of *per* mRNA<sup>22</sup> are also depicted. *b*, Comparison of immunoprecipitated  $\beta$ -galactosidase activity. Samples were roughly equal ( $\pm 10\%$ ) in total protein. Genotypes are shown of flies crossed with each other (X) or mixed after collection (+). Control experiments (not shown) with nonspecific antibodies demonstrated the specificity for the anti-HA antibody.

METHODS. Fly heads were isolated from 2–3  $\times 10^3$  entrained flies. Fly



heads (~0.3 ml) in two volumes of buffer I (20 mM HEPES, 100 mM KCl, 1 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 10% glycerol, 0.5% Triton X-100, 0.1% Nonidet-40, 1 mM DTT, pH 7.5) with protease inhibitors (1 mM PMSF, 20  $\mu$ g ml<sup>-1</sup> aprotinin, 10  $\mu$ g ml<sup>-1</sup> leupeptin, 2  $\mu$ g ml<sup>-1</sup> pepstatin, A) were homogenized at 4°C for ~20 s (Kontes mimi-homogenizer) and then sonicated briefly. The extract was spun to remove cellular debris and assayed for  $\beta$ -galactosidase<sup>27</sup>. In the SG- or NG-containing samples, comparable levels of activity (giving similar total protein) were used for immunoprecipitation with 12CA5 antibody (Fig. 1 legend). Immunoprecipitates in assay buffer<sup>27</sup> at 37°C were agitated (for 15–20 h) and the absorbance at 595 nm was corrected for the contribution from extracts of HA *per* transformants (no  $\beta$ -gal gene). Activity is expressed as ( $A_{595}$  after 1 h assay)  $\times 10^3$ .



ligands may also participate in the regulation of PER or SIM function in *Drosophila*.

Although more PAS mutants need to be analysed, the result of the PERL (V243→D243) change may explain how this affects circadian rhythms, namely by interfering with proper protein-protein interaction. The data shown in Fig. 4b indicate that these could include homotypic interactions which might be a molecular basis for the semi-dominance of most *per* mutations<sup>1,9,14</sup>. PER may also engage in heterotypic interaction with unknown protein partners, for example by sequestering one or

more bHLH-PAS proteins through a PAS-mediated protein-protein interaction, as in the mechanism proposed for the function of the HLH proteins ID (inhibitor of MyoD) and EMC (extra macrochaetae)<sup>15,16</sup>. Such a mechanism could underlie the feedback loop involving PER and the transcription of its own messenger RNA<sup>6-8</sup>, and pertain to the temporal regulation of gene expression in other organisms that is an important feature of the circadian oscillator (for example, see refs 17-21). This would be supported by the identification of a protein 'partner' that interacts with PER through its PAS domain. □

Received 23 February; accepted 28 April 1993.

1. Konopka, R. J. & Benzer, S. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2112-2116 (1971).
2. Nambu, J. R., Lewis, J. O., Wharton, K. A. & Crews, S. T. *Cell* **67**, 1157-1167 (1991).
3. Hoffman, E. C. *et al. Science* **252**, 954-958 (1991).
4. Burbach, K. M., Poland, A. & Bradfield, C. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8185-8189 (1992).
5. Bayliss, M. K., Bargiello, T. A., Jackson, F. R. & Young, M. W. *Nature* **326**, 390-392 (1987).
6. Hardin, P. E., Hall, J. C. & Rosbash, M. *Nature* **343**, 536-540 (1990).
7. Hardin, P. E., Hall, J. C. & Rosbash, M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11711-11715 (1992).
8. Liu, X. *et al. J. Neurosci.* **12**, 2735-2744 (1992).
9. Rutala, J. E., Edery, I., Hall, J. C. & Rosbash, M. *J. Neurogenet.* **8**, 101-113 (1992).
10. Zerr, D. M., Hall, J. C., Rosbash, M., & Siwicki, K. K. *J. Neurosci.* **10**, 2749-2762 (1990).
11. Blackwood, E. M. & Eisenman, R. N. *Science* **251**, 1211-1217 (1991).
12. Forman, B. M. & Samuels, H. H. *Molec. Endocrin.* **4**, 1293-1301 (1990).
13. Fawell, S. E., Lees, J. A., White, R. & Parker, M. G. *Cell* **60**, 953-962 (1990).
14. Smith, R. F. & Konopka, R. J. *Molec. gen. Genet.* **189**, 30-36 (1982).
15. Benezra, R., Davis, R. L., Lockshorn, D., Turner, D. L. & Weintraub, H. *Cell* **61**, 49-59 (1990).
16. Van Doren, M., Ellis, H. M. & Posakony, J. W. *Development* **113**, 245-255 (1991).

17. Dunlap, J. C. & Feldman, J. F. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1096-1100 (1988).
18. Raju, U., Constantinos, K., Nuneu-Regueiro, M. & Eskin, A. *Science* **253**, 673-675 (1991).
19. Rosbash, M. & Hall, J. C. *Neuron* **3**, 387-398 (1989).
20. Dunlap, J. C. *Trends Genet.* **60**, 159-165 (1990).
21. Khalsa, S. B. S., Whitmore, D. & Block, G. D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10862-10866 (1992).
22. Citri, Y. *et al. Nature* **326**, 42-47 (1987).
23. Kolodziej, P. A. & Young, R. A. *Meth. Enzym.* **194**, 508-519 (1991).
24. Schindler, U., Menkens, A. E., Beckman, H., Ecker, J. R. & Cashmore, A. R. *EMBO J.* **11**, 1261-1273 (1992).
25. Evan, G., Lewis, G., Ramsay, G. & Bishop, J. M. *Molec. cell. Biol.* **5**, 3610-3616 (1985).
26. Liu, X., Lorenz, L. Y., Q., Hall, J. C. & Rosbash, M. *Genes Dev.* **2**, 228-238 (1988).
27. Zwiebel, L. J., Hardin, P. E., Liu, X., Hall, J. C. & Rosbash, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3882-3886 (1991).
28. Ema, M. *et al. Biochem. biophys. Res. Commun.* **184**, 246-253 (1992).

ACKNOWLEDGEMENTS. We thank K. Curtin and K. Kreutter for help with site-directed mutagenesis; S. Crews and O. Hankinson for cDNAs encoding SIM and ARNT, respectively; H.V. Colot, J. C. Hall and R. Sen for comments on the manuscript; and T. Tishman for secretarial assistance. The work was supported by a grant from the NIH to M.R.

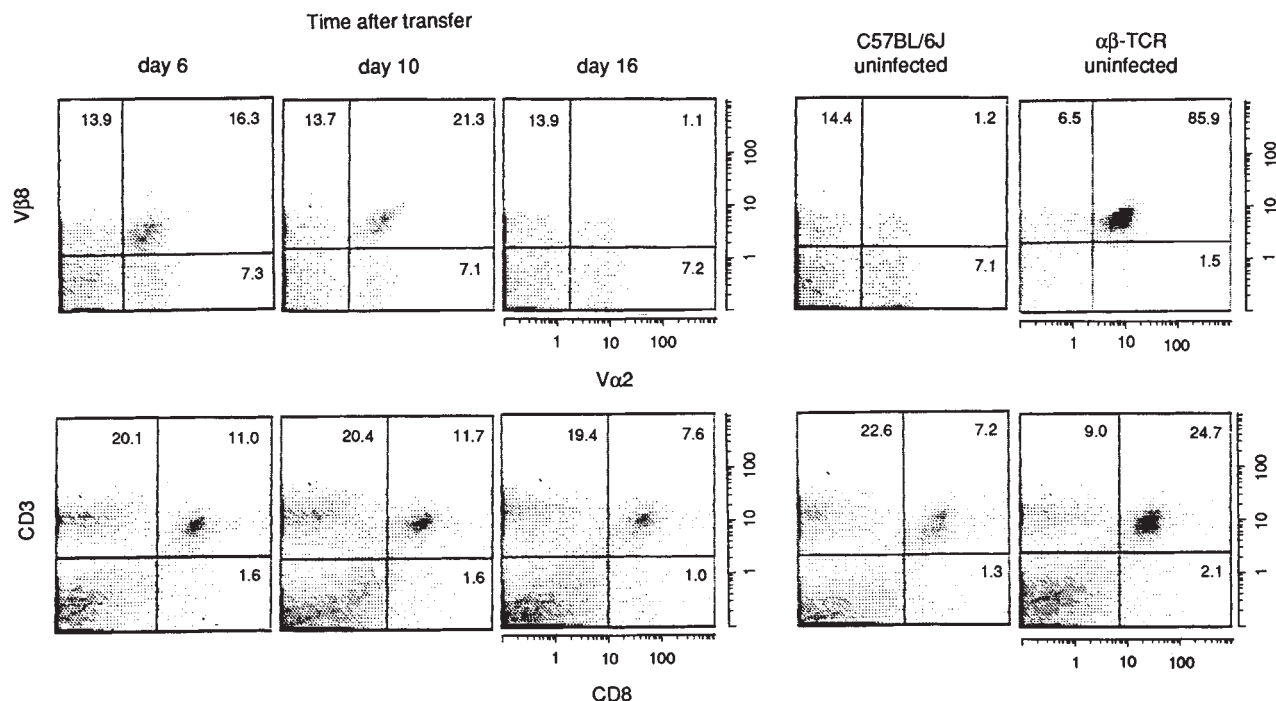
## ERRATUM

### Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells

Demetrius Moskophidis, Franziska Lechner, Hanspeter Pircher & Rolf M. Zinkernagel

*Nature* **362**, 758-761 (1993)

DURING the production process the top and bottom panels of Fig. 3 in this letter were accidentally switched. The correct figure is shown below; the legend remains the same.



# Scientific mixed bag for midsummer

Technical calculators, an automated colony picking robot, a new thermostable DNA polymerase, a cell culture system for smooth muscle cells and a software program for drawing family trees are included in this week's line-up.

HEWLETT-Packard says that its two new **technical calculators**, the HP 48GX and HP 48G, are faster and offer more functions than their predecessors, the HP 48SX and 48S (*Reader Service No. 100*). Both of the new calculators have built-in functionality, such as three-dimensional (3-D) plotting, which previously was unavailable in hand-held calculators.

The US\$350 HP 48GX has 128 Kbytes of random-access memory (RAM) and accepts plug-in cards containing memory or programs that expand upon its basic functions. The US\$165 HP 48G, with 32 Kbytes of RAM and all the features of the HP 48GX, except the expansion capability, is designed



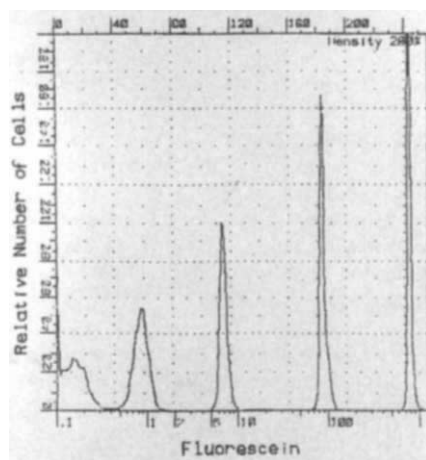
**HP 48GX technical calculator.**

for users who have no need of a memory-expansion application card. New features that enhance the HP 48GX's graphics include a simplified form-driven user interface, shading, 3-D plots, trace, shading and simultaneous plots. The HP 48GX integrates calculus and graphics functions to find roots, intersections, local extremes, derivatives, slopes and areas under curves. It displays graphs in 15 formats. Other features include 'fill-in-the-blanks' input forms, dialogue boxes for easy access to functions, and built-in scientific and engineering equations. Customers who own the HP 48SX or the HP 48S can upgrade to the HP 48GX for US\$195 or the HP 48G for US\$95.

Micro-Bridges from Crystal micro-systems have been designed to carry out **sitting drop crystallizations** in standard tissue culture plates (*Reader Service No. 101*). They are made of polystyrene and transparent. In addition, the company points out that the surface of the indentation is highly polished to avoid any undesired nucleation sites and to facilitate the visual inspection of the drops under the microscope. Micro-Bridges are at present available with a round or a square indentation. The maximum drop volume is 35  $\mu$ l for the round; 45  $\mu$ l for the square variety. One special feature of Micro-Bridges is that they fit into the wells of standard tissue culture plates (the well diameter should be approximately 16 mm). Once placed inside the wells,

the company claims that Micro-Bridges are stable and there is no need to stick them to the wells.

Spherotech has introduced the Sphero Rainbow Calibration Particles for the routine **calibration of flow cytometers** (*Reader Service No. 102*). Having uniform size, the Rainbow Calibration Particles consist of particles with four or five different fluorescence intensities. The same set of particles may be used to calibrate all available chan-



**Rainbow calibration particles (7  $\mu$ m) for calibrating flow cytometers.**

nels in flow cytometers. Spherotech says that the particles, which are available in particle sizes of 2.0  $\mu$ m, 3.5  $\mu$ m and 7.0  $\mu$ m, are very stable and packaged in a convenient dropper bottle to facilitate dispensing.

CID, through Cottenham Instruments and Materials is introducing the CI-203 **portable laser area meter**, which enables researchers to carry out nondestructive plant leaf area measurements in the field or laboratory (*Reader Service No. 103*). These measurements can be important in, for example, plant growth trials, when leaf area measurements are often carried out on the

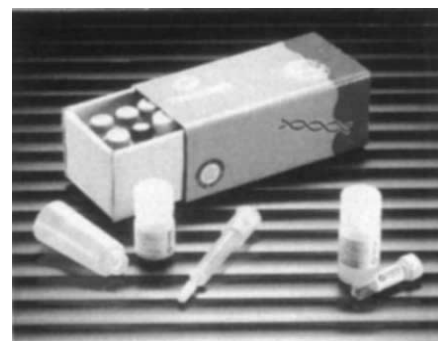


**Hand-held laser area meter for plant science.**

same leaf throughout the life cycle of the plant. The high-resolution laser scanner, microprocessor, data logger and display are all enclosed in a single hand-held unit weighing only 1 kg. The company says that precise measurements can also be performed on tender leaves, leaves with irregular margins, and diseased and insect damaged leaves; a transparent sheath is provided for extensive tissue damage. All six measurement parameters—area, width, length, perimeter, shape factor and aspect ratio—comprise a data set and are compiled at the same time. Over 16,000 data entries can be stored, or they can be transferred to a computer or printer. The CI-203 is compatible with the CI-203CA conveyor attachment, which is designed for measuring multiple detached leaves.

## ■ Molecular miscellany

Promega's new PinPoint Xa **protein purification system** (for research use only) is designed to be a simple, reliable non-denaturing method of expressing and purifying recombinant fusion proteins in *Escherichia coli* (*Reader Service No. 104*). PinPoint Xa vectors contain a sequence that becomes biotinylated in *E. coli*. The biotinylated protein binds to SoftLink Resin, which becomes functional with monomeric avidin. The bound protein can be released under non-denaturing conditions (that is, 5 mM biotin). A Factor Xa protease cleavage



**Protein purification Promega's way.**

site is located between the biotinylated peptide and target protein, allowing cleavage and purification of the protein of interest. Promega says that proteins can be purified using the PinPoint column provided or in batch mode. The fusion protein can be monitored throughout the purification process using the streptavidin alkaline phosphatase provided. The complete system includes the PinPoint Xa-1/2/3 vectors (three reading frames), PinPoint control vector, the SoftLink resin, streptavidin alkaline phosphatase and the PinPoint purification column.

**\$3.50/Base**

**Custom DNA  
Synthesis**

No Set-up Fee,  
Shipped in 24 hours.

**1-800-788-4362  
1-800-786-4362 Fax**

**KEYSTONE  
Laboratories, Inc.**

Reader Service No.33

**Highest Quality  
Oligonucleotides**

Synthesised, purified and shipped in 3 days

**£2.75 per base**

(All inclusive)

0.2 µM scale (approx. 500 µg)

- Antisense (completely or partially sulphurised)
- Extensive range of labelling and modifications
- Full design and analysis facilities
- Gene cassettes, bulk order and contract discounts

Molecular Medicine Unit

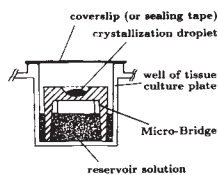
King's College School of Medicine & Dentistry

Tel. 071 - 326 3126 FAX 071 - 733 3877

Reader Service No.56

**Micro-Bridges**

for sitting drop crystallizations



For further information  
and free samples contact:

**Crystal microsystems**

47 Purcell Road  
OXFORD, OX3 0HB, U.K.

FAX:(UK)-865-200327

Reader Service No.7

YYYYYYYYYYYYYYYYYYYY

**Custom  
Monoclonal Antibody  
Secreting Hybrids**

**MONOCLONAL  
IMMUNOGLOBULINS MIG**

Tel: 800-995-6447

209-754-0136

Fax: 209-754-0225

YYYYYYYYYYYYYYYYYYYY

Reader Service No.4

The FIGE Mapper Electrophoresis System from Bio-Rad utilizes **field inversion gel electrophoresis**, or FIGE, to provide high resolution separations of all electrophoretic techniques for DNA molecules ranging from 100 to 150,000 base pairs (*Reader Service No. 105*). The device contains ten stored programmes that are designed to cover the size range of 1 kb to 150 kb; run conditions can also be preprogrammed manually and up to ten programmes can be stored in nonvolatile memory. Accommodating both asymmetric pulse time and asymmetric voltage formats, the FIGE Mapper can generate nonlinear shaped pulse time ramps to enhance the separation of DNA molecules. Bio-Rad says that this capability yields optimal separations of the cloned inserts from M13, lambda, cosmid, and P1 phage vectors, bacterial artificial chromosomes, genomic DNA for Southern hybridization analysis, or DNA fingerprinting for forensic analysis. The complete system includes a power module (power supply, switcher and electrode drivers), horizontal electrophoresis cell, casting tray, accessories and a variable speed pump. The pump recirculates electrophoresis buffer to dissipate heat, allowing separations to be performed at room temperature, thus eliminating the requirement for buffer chillers (which can be expensive).

Picking colonies by hand may be a thing of the past with the new **automated colony picking and gridding robot** developed by Hybaid (*Reader Service No. 106*). Under Apple Macintosh computer control, the new Colony Picker uses image processing technology to identify and pick randomly plated libraries of *Escherichia coli*, yeast and plaques, including all commonly used colour-based selection systems. Once loaded with plates, Hybaid says that the robotic device will pick and order 1,000 colonies per hour for over 8 hours in a single unattended session. The device is also capable of automatically constructing gridded arrays onto nylon filters up to 20 × 20 cm in size, at a range of densities up to 25 × 96 well plates. Additional software features allow the user to grid and screen up to 150,000 clones on a single 20 × 20 cm filter.

Perkin-Elmer has introduced a new DNA polymerase called UITma, which is a **thermostable DNA polymerase** designed specifically to correct polymerase chain reaction (PCR) misincorporations and prevent misextensions while maintaining amplification efficiency (*Reader Service No. 107*). Created for high fidelity PCR cloning applications, the company says that the enzyme has been genetically engineered to achieve an optimal balance between proofreading and polymerase activities and is guaranteed for PCR performance. Derived from *Thermotoga maritima*, UITma DNA polymerase is optimized to repair 3'-mismatches in PCR amplification, provide a high yield

of specific PCR products, and produce blunt-ended PCR products that are suitable for cloning and gene expression.

**Neuroscience news**

Human tenascin — a **disulphide-linked glycoprotein** with three subunits of 190, 200 and 220 kilodaltons — is now available from Chemicon in a purified form (*Reader Service No. 108*). Tenascin is expressed at specific times during development of the central nervous system and is thought to mediate neural crest cell migration and glia-neuron interactions. It also appears in a variety of tissues, often in basement membranes or intercellular spaces. In adults, tenascin is expressed during onco-development, usually in the vasculature of tumours and in the tumour extracellular matrix. Also available are Chemicon's monoclonal and polyclonal antibodies for tenascin research.

Normal human aortic smooth muscle cells are now being made available by Clonetics for use in basic and applied research in vascular pathology, including atherosclerosis and cardiovascular pharmaceutical development. They are available in the MyoPack-AOSMC, a complete normal **human aortic smooth muscle cell culture system** containing either cryopreserved or proliferating



**Culture system for smooth muscle cells.**

cells (*Reader Service No. 109*). Packs also include optimized low serum (5 per cent) smooth muscle cell growth medium and all the reagents necessary for propagation and subculture, *in vitro*. All items contained in the pack may be purchased individually. In addition to the general field of cardiovascular research, stated applications include ion exchange research, signal transduction studies and adhesion assays.

Axon Instruments has released the Axoclamp-2B, the new **intracellular microelectrode recording amplifier** with high compliance for recording from large cells (*Reader Service No. 110*). Like its predecessor, the Axoclamp-2A, the new model is intended for a variety of electrophysiological applications involving intracellular microelectrode recordings. With two independent amplifiers, the US\$7,700



Axoclamp-2B supports both continuous and discontinuous single-electrode current and voltage clamping. The company says that the versatility of the amplifier has been enhanced with a new, high-compliance, two-electrode voltage-clamp mode, which makes the device suitable for recording the large currents of *Xenopus* oocytes. A variety of optional headstages is offered, enabling the Axoclamp-2B to be used in several recording configurations. The new 'bath-clamp' headstage can be used to clamp the potential of the bath to zero, thereby eliminating the voltage error due to current flowing through the bath-grounding electrode. Axon Instruments says that this configuration also compensates for changes in bath potential resulting from alterations in the temperature and

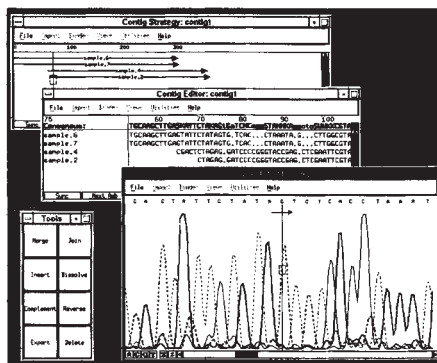


**Axon Instruments' Axoclamp-2B intracellular microelectrode recording amplifier.**

ionic composition of the cell's bathing solution. Built-in features include: precision command current and voltage sources, a preset command to clamp the cell's resting membrane potential, offsets, meters and filters. The Axoclamp-2B can also clear microelectrodes and assist in cell penetration.

### ■ Assorted software

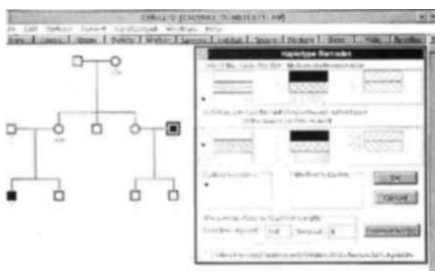
IntelliGenetics has launched GEneration, a new software program for large-scale sequence assembly and management that uses a multiple sequence alignment algorithm that has been developed especially for fragment assembly (*Reader Service No. 111*). The algorithm, the company says, provides highly sensitive merges and eliminates degradation of the consensus sequence or 'consensus rot', which can occur with other sequence assembly software packages. In



**Large-scale sequence assembly and management program for the Sun.**

2 min 22 s, GEneration, which runs on a Sun SPARCstation, assembled 114 fragments, totalling over 36,500 base pairs, into a 9,351 base pair consensus sequence. In addition, GEneration has also successfully assembled a test project into a final consensus sequence of 2,000,000 base pairs. The software keeps a history of the sequencing project, displays the merge strategy and plots the redundancy level of the final consensus sequence. It also screens sequence fragments for repeating sequences, vectors and enzymes.

Cherwell Scientific announces the release of Cyrillic, a Windows-based pedigree drawing software program for geneticists (*Reader Service No. 112*). Using Cyrillic, users can import pedigrees and marker data from standard sources such as MLINK pedigree files. Pedigree data are displayed graphically, and there is complete



**Cyrillic pedigree drawing package.**

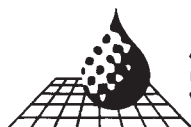
control over line widths, fonts, colours, symbol sizes, scaling of the drawing and the information displayed about each symbol. The drawings can be exported to word processor documents using the Windows clipboard. Cyrillic runs on any IBM PC or compatible computer running Microsoft Windows 3.1 and costs US\$600.

Oxford Molecular, a molecular modelling software company, has introduced a new tool for improving three dimensional quantitative structure activity relationships (QSAR) techniques (*Reader Service No. 113*). Tsar 2.0 has a new interface to its similarity software, Asp 2.0, which gives researchers access to a new technique explored at Graham Richards' laboratory at the University of Oxford. The technique combines molecular similarity calculations with QSAR studies. The combined Tsar-Asp system provides an automatic method of optimizing molecular alignment and conformation of candidate compounds relative to lead compounds (those already exhibiting desired activity/properties). An index is calculated to give a quantitative measure of the similarity between any two molecules. Oxford Molecular points out that with Tsar-Asp, this index can be calculated using a choice of electronic, lipophilic or steric interactions. And as these are the forces that cause molecules to be attracted to and oriented in a receptor site, the 3-D data included in the QSAR analysis is more likely to correlate with activity. Tsar 2.0 and Asp 2.0 are available on Silicon Graphics, IBM RS/6000 and Hewlett-Packard UNIX workstations.

### ■ Kits and reagents

Miltenyi Biotec now offers the new CD34 progenitor cell isolation kit, which uses magnetic separation technology to separate CD34 expressing cells from peripheral blood, cord blood and bone marrow, with high purity and high rates of recovery (*Reader Service No. 114*). When using the kit, the company says that up to 10,000fold enrichment can be achieved in less than 1 hour, that no bead detachment is required as MACS

ADVERTISEMENT



### Severn Biotech Ltd

Unit 23, Hoobrook Enterprise Centre,  
Worcester Road, Kidderminster DY10 1HY

#### Custom DNA & RNA Oligo's

Probes, Primers, Antisense Oligo's, Ribonucleotides, 5' and 3' labelling with Biotin Fluorescein, Rhodamine, Modified bases purification etc.

#### Complete Peptide Service

Small Scale Multiple Peptides 5mgs or larger scale 25mgs - 100mgs using L and D Amino Acids, screening grade or HPLC purified. Phosphorylated, Acetylated Amidated and Protein Carrier Conjugated.

**For a Fast Efficient Service Reply Today**

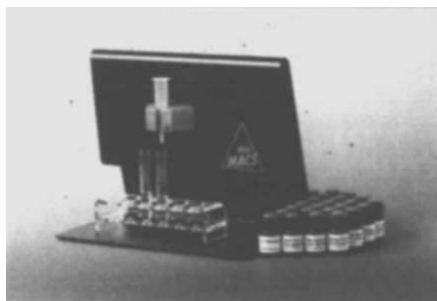
**Tel: (0562) 825286 Fax: (0562) 825284**

### CUSTOM SERVICES

**Comprehensive Custom Services for DNA, RNA and Peptides from SBL**

- DNA Synthesis
- RNA Synthesis
- Multiple Peptide Synthesis (5mgs)
- Larger Scale Peptide Synthesis 25 - 100 mgs
- HPLC Purification
- Custom Conjugation to Carriers
- Peptide Sequencing

Reader Service No.21



**CD34 magnetic cell isolation kit.**

microbeads are one million times smaller than a cell and biodegradable, and that the bound cells have virtually no effect on cell function or viability. Cells isolated by this procedure can be used directly in flow cytometry, functional experiments or cell culture.

Fluka has introduced the MicroSelect line of **high purity polyethylene glycol** types to suit most biochemical applications (Reader Service No. 115). These specially selected and purified products have been designed to meet rigorous quality control requirements: they are free of impurities such as peroxides, aldehydes or metal ions, contain no insoluble matter, and show only limited absorbance at 260 nm and 280 nm. Polyethylene glycol for use in molecular biology applications is also available in the new quality grade. These products are in addition free of DNases, RNases, proteases and phosphatases.

TDA Research has begun **commercial production of fullerenes** using a combustion synthesis process. These materials are available through Research Materials (Reader Service No. 116). The company says that the combustion process has several important advantages over the Krätschmer-Huffman carbon arc process generally used to produce fullerenes: it can be used to produce a much higher fraction of  $C_{70}$  than the arc process; the  $C_{70}$  fraction in the product can be varied over a wide range by changing the operating conditions; the process is continuous, uses inexpensive and convenient feedstocks (liquid hydrocarbons and oxygen), and is easily scaled up; and it is easy to control and produces a product with a very uniform composition. The combustion synthesis process uses a low pressure, premixed benzene/oxygen flame to produce a mixture of fullerenes ( $C_{60}$ ,  $C_{70}$  and higher). Although fullerenes are produced over a wide range of conditions, the process is typically run at carbon/oxygen ratios greater than 0.8 (equivalence ratios greater than 2), pressures of 20 to 50 torr and temperatures in excess of 2,000 K.

## Picture perfect

With the new MultiMode **scanning probe microscope** from Digital Instruments, users can tailor the scanning probe microscopy

technique to the specific characteristics of their sample (Reader Service No. 117). This new microscope supports atomic force microscopy and scanning tunnelling microscopy in conventional and electrochemical operation. Samples can be scanned in several atomic force microscope modes, including contact, noncontact, magnetic force, and the new Tapping mode, which, the company says, allows soft samples to be imaged without damage by eliminating lateral compressive and shear forces.

Last month, Polaroid introduced the MicroCam SLR automatic, **single-lens-reflex instant camera** for photographing specimens through a light microscope (Reader Service No. 118). The US\$795 MicroCam SLR, which can be used to produce 'instant' colour or black and white prints at the push of a button, will attach to most light microscopes by sliding its built-in  $\times 10$  magnification lens directly into the microscope's eyepiece tube or phototube. Polaroid says that MicroCam's multi-element glass lens produces sharp, clear pictures, and the bright, single-lens-reflex viewing system makes it easy for the microscopist to focus and frame the specimen. The LCD panel reports all camera functions in any of six user-selectable languages — English, French, German, Italian, Spanish and Japanese. The display shows



**Polaroid's MicroCam SLR camera features microprocessor-controlled exposure and filtration, digital light measurement and a rotary shutter.**

the proposed exposure and offers the user the option of terminating the exposure before it begins. The electronic shutter release and the two-second exposure delay are designed to ensure that there is no vibration during exposure. The camera's photometer and microprocessor calculate any exposure between 1/60th of a second and 16 min. At the end of each exposure, the camera automatically ejects the film, which develops without the need to time or peel the print. A manual version is also available.

Imaging Services, an independent laboratory that specializes in scanning probe

microscopy, now offers **microscopy services** using Digital Instruments' NanoScope III MultiMode scanning probe microscopes (Reader Service No. 119). The MultiMode atomic force microscope offers TMAFM, which can be used to produce images of soft biological samples, non-destructively and with high resolution. Imaging Services also offers use of a prototype fluorescence optical microscope with concurrent atomic force microscopy, which should be of particular interest to biologists. Samples can be imaged in a variety of fluids, including water, phosphate buffered saline and propanol with no coating or special preparation required. □

*These notes are compiled by Diane Gershon from information provided by the manufacturers. For more information, fill in the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.*

## ADVERTISEMENTS

### Teflon\* Coated and Bare Wire from MEDWIRE



Medwire provides the highest quality platinum, platinum alloy, silver and stainless steel wire in both single and multistrand configurations.

For prompt delivery or additional information, please contact:

**MEDWIRE**, an affiliate of Sigmund Cohn Corp.  
121 So. Columbus Ave. Mount Vernon, NY 10553  
Tel: (914) 664-5300 Fax: (914) 664-5377

\*Reg. TM of E. I. DuPont & Co

Reader Service No.25

### Plant and Animal Tissue HOMOGENIZER and EMULSIFIER

- Disrupts and disperses small samples (1/2-100 ml)
- Both mechanical and ultrasonic action
- 5,000-30,000 rpm
- Built-in speed control
- Easy to clean
- Best price (\$485 complete)



**The TISSUE-TEAROR™**  
Manufactured By:  
**BIOSEPEC PRODUCTS** • PH/FAX (918) 336-3363  
P.O. BOX 722 • BARTLESVILLE, OKLAHOMA 74005

Reader Service No. 15